EXHIBIT C
Part 3 of 3

The Washington Post
Democracy Dies in Darkness

Here's how to keep Russian hackers from attacking the 2018 elections

By **J. Alex Halderman** and **Justin Talbot-Zorn**

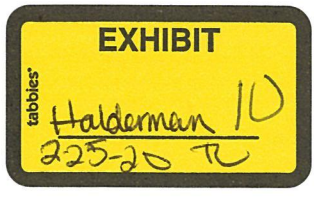
June 21, 2017 at 6:00 a.m. EDT

"They're coming after America," former FBI director James B. Comey told the Senate intelligence committee this month. "They will be back."

In a highly politicized hearing, this bold statement drew strikingly little partisan disagreement. Senators on both sides of the aisle have seemingly reached consensus that foreign agents did try to tamper with the 2016 election and that they are extremely likely to do so again.

The question is: What do we do about it?

While the ongoing Russia investigation has, understandably, received massive attention, there's so far been scant public focus on the question of how we safeguard our electoral systems from outside interference in the future. Responding to the threat of election hacking isn't exclusively a matter of diplomatic intrigue or international sanctions. It's fundamentally a matter of computer science: how we harden our election technology through cybersecurity standards.



This week, we're joining a group of more than 100 experts on election administration, computer science and national security in releasing a letter that lays out an actionable plan for safeguarding the vote. The experts include tea party Republicans and progressive Democrats, academic computer scientists and corporate security officials — all united in the view that our nation's rough patchwork of voting security measures is wholly inadequate. One of us (Halderman) will testify Wednesday before the Senate Intelligence Committee on Russia's attacks last year.

This shouldn't be news to lawmakers. In the past decade, cybersecurity experts have revealed devastating vulnerabilities in every U.S. voting machine they've studied. In 2014, the bipartisan Presidential Commission on Election Administration sounded the alarm about an "impending crisis" of insecure voting technology. In 2015, Lawrence Norden and Christopher Famighetti of the Brennan Center for Justice at New York University showed in a comprehensive study that the nation's voting machines are largely past their shelf-lives and deeply insecure. According to a survey of 274 election administrators across 28 states, a strong majority of election officials claim they need security upgrades to voting machines but simply lack the resources.

Ten years ago, Halderman was part of the first academic research team to conduct a comprehensive security analysis of a Direct Recording Electronic (DRE) voting machine. The study's findings were deeply troubling: It's possible to reprogram a machine to cause any candidate to win, without leaving a trace. The research team created malicious software — vote-stealing code — that could spread from machine to machine, much like a computer virus, and invisibly change the election outcome. Since then, cybersecurity experts have studied a wide range of U.S. voting machines — including both touch screens and optical scanners — and in *every single case*, they found severe vulnerabilities that would allow attackers to sabotage machines or alter votes

This month's blockbuster reporting in the Intercept and Bloomberg News show that hostile nations have our computerized election infrastructure in their sights. And the threats aren't limited to the voting machines and tabulators: adversaries can also go after voter registration databases and electronic poll books to block voters, create long lines at polling places and instill distrust in the system.

So why hasn't Congress acted?

One simple answer is that lawmakers need a straightforward policy agenda to fix the system. The new statement from the 100 election security experts provides a concrete road map:

First, Congress should provide time-sensitive matching funds to states to upgrade voting technologies, and, in particular, replace paperless DRE voting machines with systems that include a good old-fashioned paper ballot — that is to say, a physical record of the vote that's out of reach from cyberattacks.

Second, Congress should call on states to conduct risk-limiting audits for every federal race, by inspecting enough of the paper ballots to tell whether the computer results are accurate. These audits are a common-sense quality control, and they should be routine. Since they only require officials to check a small random sample of ballots, they quickly and affordably provide high assurance that the election outcome was correct. As Ron Rivest of the Massachusetts Institute of Technology and Philip Stark of the University of California have explained, states can gain high confidence regarding election outcomes by checking as few as 0.5 percent of the ballots in a given contest.

Finally, Congress should instruct federal agencies to partner with states to conduct serious and comprehensive threat assessment, and to identify and apply best practices in cybersecurity from across sectors to the design of voting equipment and the management of federal elections. This will raise the bar for attacks of all sorts.

There's evidence this agenda can fly even in the age of hyperpartisan gridlock.

While many Democrats have supported election security reforms since former Rep. Rush Holt proposed related reforms a decade ago, prominent conservatives are now championing the cause. Recently, retired Army Intelligence Lt. Col. Tony Shaffer — a Fox News contributor and fearless President Barack Obama critic — joined former CIA director R. James Woolsey — a leading national defense advocate — to call for audits and federal cybersecurity standards. In a Fox News op-ed last month, the two made a conservative case for election security reform as a matter of national security, explaining why, among other factors, Congress' unfunded mandates under the Help America Vote Act of 2002 justify new security investments. Shaffer and Woolsey quote President Trump himself from an interview the morning of the election: "There's something really nice about the old paper ballot system," the then-candidate states. "You don't worry about hacking."

Perhaps the strongest argument why the new federal election security agenda can succeed is cost. New analysis from the Brennan Center finds that the country can replace insecure paperless voting systems for somewhere between $130 million and $400 million. Implementing risk-limiting audits nationally for federal elections would cost less than $20 million a year. These amounts are a rounding error in the administration's $640 billion defense budget request, but the investment would be a guaranteed way to boost voter confidence and significantly strengthen an important element of our national security.

With many state and local officials keen to make necessary tech upgrades, Congress may need to only cover a fraction of the overall costs.

If lawmakers agree with Comey's assessment that foreign agents are "coming after America," it stands to reason that Congress should devote resources to addressing the threat. This is a small price tag for the defense of our democracy.

J. Alex Halderman
J. Alex Halderman is professor of computer science at the University of Michigan and director of Michigan's Center for Computer Security and Society. Follow

Justin Talbot-Zorn
Justin Talbot-Zorn is a Truman National Security Fellow and an adviser to the National Election Defense Coalition. He has served as legislative director to three members of Congress. Follow

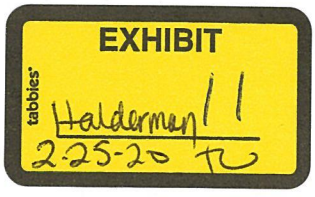
Want to Know if the Election was Hacked? Look at the Ballots

J. Alex Halderman Follow
Nov 23, 2016 · 7 min read

You may have read at NYMag that I've been in discussions with the Clinton campaign about whether it might wish to seek recounts in critical states. That article, which includes somebody else's description of my views, incorrectly describes the reasons manually checking ballots is an essential security safeguard (and includes some incorrect numbers, to boot). Let me set the record straight about what I and other leading election security experts have actually been saying to the campaign and everyone else who's willing to listen.

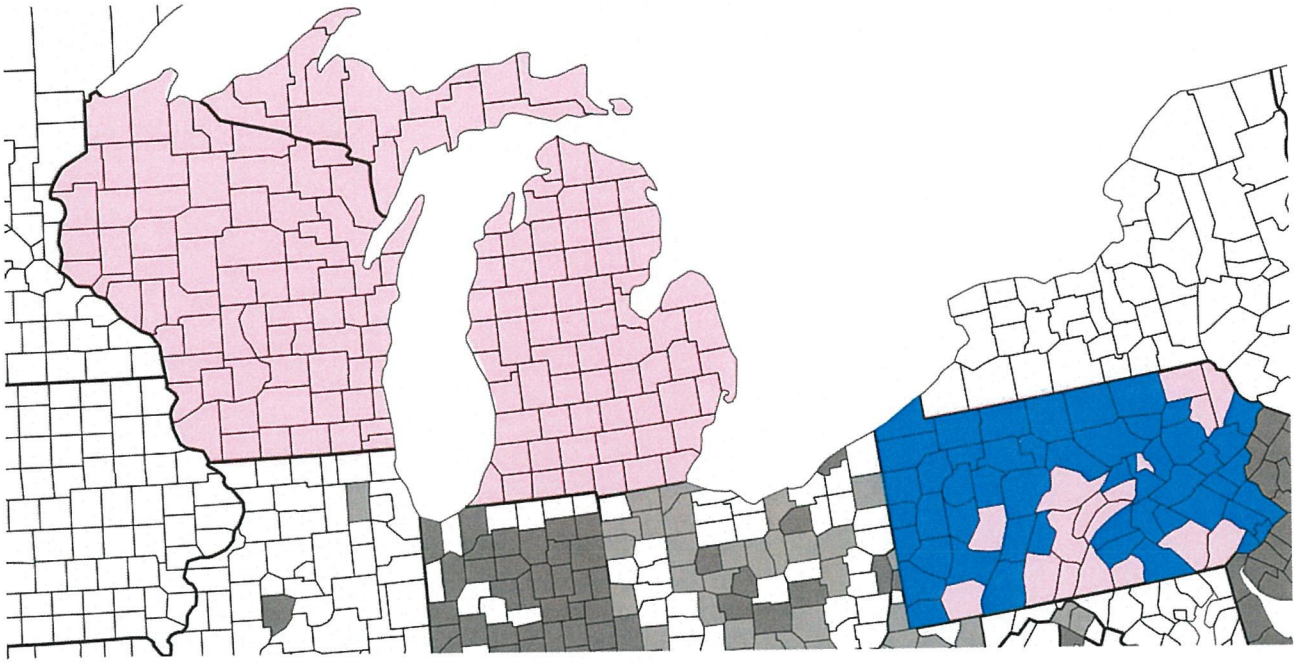
How might a foreign government hack America's voting machines to change the outcome of a presidential election? Here's one possible scenario. First, the attackers would probe election offices well in advance in order to find ways to break into their computers. Closer to the election, when it was clear from polling data which states would have close electoral margins, the attackers might spread malware into voting machines in some of these states, rigging the machines to shift a few percent of the vote to favor their desired candidate. This malware would likely be designed to remain inactive during pre-election tests, do its dirty business during the election, then erase itself when the polls close. A skilled attacker's work might leave no visible signs — though the country might be surprised when results in several close states were off from pre-election polls.

Could anyone be brazen enough to try such an attack? A few years ago, I might have said that sounds like science fiction, but 2016 has seen unprecedented cyberattacks aimed at interfering with the election. This summer, attackers broke into the email system of the Democratic National Committee and, separately, into the email account of John Podesta, Hillary Clinton's campaign chairman, and leaked private messages. Attackers infiltrated the voter registration systems of two states, Illinois and Arizona, and stole voter data.



And there's evidence that hackers attempted to breach election offices in several other states.

In all these cases, Federal agencies publicly asserted that senior officials in the Russian government commissioned these attacks. Russia has sophisticated cyber-offensive capabilities, and has shown a willingness to use them to hack elections. In 2014, during the presidential election in Ukraine, attackers linked to Russia sabotaged the country's vote-counting infrastructure and, according to published reports, Ukrainian officials succeeded only at the last minute in defusing vote-stealing malware that was primed to cause the wrong winner to be announced. Russia is not the only country with the ability to pull off such an attack on American systems — most of the world's military powers now have sophisticated cyberwarfare capabilities.



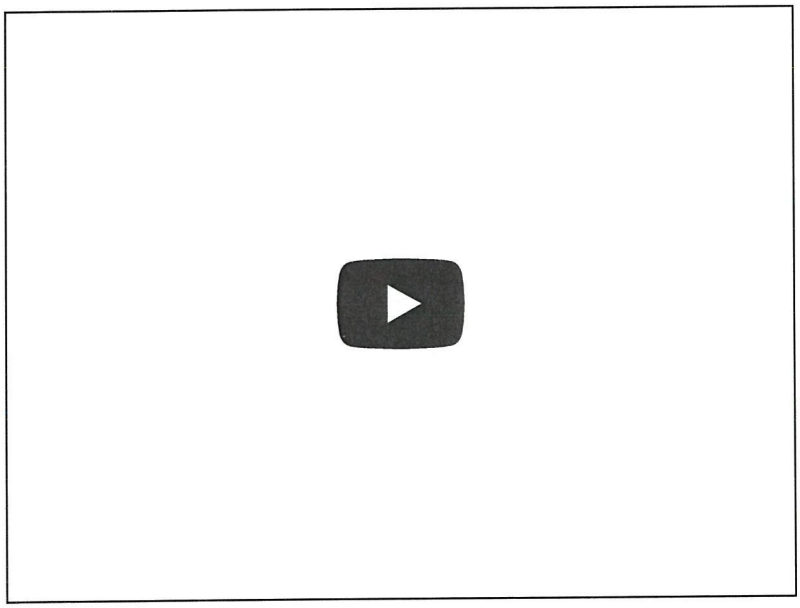
The pink counties predominately use optical scan paper ballots, which can be examined to confirm that the computer voting machines produced an accurate count. Blue counties use paperless voting systems, which require forensic analysis.

Were this year's deviations from pre-election polls the results of a cyberattack? Probably not. I believe the most likely explanation is that the polls were systematically wrong, rather than that the election was hacked. But I don't believe that either one of these seemingly unlikely explanations is overwhelmingly more likely than the other. The only way to know whether a cyberattack changed the result is to closely examine the available physical evidence — paper ballots and voting equipment in critical states like

Wisconsin, Michigan, and Pennsylvania. Unfortunately, **nobody is ever going to examine that evidence unless candidates in those states act now, in the next several days, to petition for recounts.**

What's to stop an attack like this from succeeding?

America's voting machines have serious cybersecurity problems. That isn't news. It's been documented beyond any doubt over the last decade in numerous peer-reviewed papers and state-sponsored studies by me and by other computer security experts. We've been pointing out for years that voting machines are computers, and they have reprogrammable software, so if attackers can modify that software by infecting the machines with malware, they can cause the machines to give any answer whatsoever. I've demonstrated this in the laboratory with real voting machines — in just a few seconds, anyone can install vote-stealing malware on those machines that silently alters the electronic records of every vote.



It doesn't matter whether the voting machines are connected to the Internet. Shortly before each election, poll workers copy the ballot design from a regular desktop computer in a government office, and use removable media (like the memory card from a digital camera) to load the ballot onto each machine. That initial computer is almost certainly not well secured, and if an attacker infects it, vote-stealing malware can hitch a ride to every voting machine in the area. There's no question that this is possible for technically sophisticated attackers. (If my Ph.D. students and I were criminals, I'm sure we could pull it off.) If anyone reasonably skilled is sufficiently motivated and willing to face the risk of getting caught, it's happened already.

Why hasn't more been done about this? In the U.S., each state (and often individual counties or municipalities) selects its own election technology, and some states have taken steps to guard against these problems. (For instance, California banned the use of the most dangerous computer voting machines in 2007 as a result of vulnerabilities that I and other computer scientists found.) But many states continue to use machines that are known to be insecure — sometimes with software that is a decade or more out of date — because they simply don't have the money to replace those machines.

There is one absolutely essential security safeguard that protects most Americans' votes: paper.

I know I may sound like a Luddite for saying so, but most election security experts are with me on this: *paper ballots are the best available technology for casting votes.* We use two main kinds of paper systems in different parts of the U.S. Either voters fill out a ballot paper that gets scanned into a computer for counting (optical scan voting), or they vote on a computer that counts the vote and prints a record on a piece of paper (called a voter-verifiable paper audit trail). Either way, the paper creates a record of the vote that can't be later modified by any bugs, misconfiguration, or malicious software that might have infected the machines.

After the election, human beings can examine the paper to make sure the results from the voting machines accurately determined who won. Just as you want the brakes in your car to keep working even if the car's computer goes haywire, accurate vote counts must remain available even if the machines are malfunctioning or attacked. In both cases, common sense tells us we need some kind of physical backup system. I and other

election security experts have been advocating for paper ballots for years, and today, about 70% of American voters live in jurisdictions that keep a paper record of every vote.

Washington Journal: J. Alex Halderman on Cybersecurity and Voting

Oct. 4, 2016: Prof. J. Alex Halderman, who has conducted research on voting machine security, talks about his concerns regarding the nation's voting...

www.c-span.org

There's just one problem, and it might come as a surprise even to many security experts: **no state is planning to actually check the paper** in a way that would reliably detect that the computer-based outcome was wrong. About half the states have no laws that require a manual examination of paper ballots, and most other states perform only superficial spot checks. If nobody looks at the paper, it might as well not be there. A clever attacker would exploit this.

There's still one way that some of this year's paper ballots could be examined. In many states, candidates can petition for a recount. The candidate needs to pay the cost, which can run into millions of dollars. The deadlines for filing recount petitions are soon — for example, this Friday in Wisconsin (margin 0.7%), Monday in Pennsylvania (margin 1.2%), and the following Wednesday in Michigan (margin 0.3%).

Examining the physical evidence in these states — even if it finds nothing amiss — will help allay doubt and give voters justified confidence that the results are accurate. It will also set a precedent for routinely examining paper ballots, which will provide an important deterrent against cyberattacks on future elections. Recounting the ballots now can only lead to strengthened electoral integrity, but the window for candidates to act is closing fast.

Much more needs to be done to secure America's elections, and important new safeguards could be put in place by 2018. States still using paperless voting machines should replace them with optical scan systems, and all states should update their audit and recount procedures. There are fast and inexpensive ways to verify (or correct) computer voting results using a risk-limiting audit, a statistical method that involves manually inspecting randomly selected paper ballots. Officials need to begin

preparing soon to make sure all of these improvements are ready before the next big election.

. . .

J. Alex Halderman is Professor of Computer Science & Engineering at the University of Michigan and Director of Michigan's Center for Computer Security & Society. His course on election technology, Securing Digital Democracy, is available on Coursera. He was recently named by Popular Science as one of the "ten brightest minds reshaping science, engineering, and the world."

Politics Elections Security Voting

Photo by Bob Foran

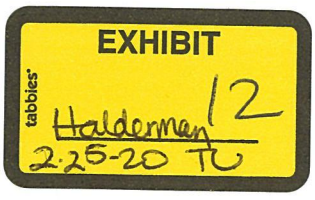


Photo by Bob Foran

HACKING THE VOTE: IT'S EASIER THAN YOU THINK

By Steve Friess | Fall 2018

PROFESSOR J. ALEX HALDERMAN HAS MADE A CAREER STUDYING ELECTRONIC VOTING SECURITY. HIS RESEARCH HAS CHANGED THE CONCEPT OF STOLEN ELECTIONS FROM THEORY TO REALITY.

"I know America's voting machines are vulnerable," J. Alex Halderman firmly stated, pausing to lift his head from the page he read to look up at a phalanx of U.S. senators, "because my colleagues and I have hacked them—repeatedly—as part of a decade of research studying the technology that operates elections and learning how to make it stronger."

It's not hyperbole to say a shudder swept through that august meeting room in the Hart Senate Office Building in Washington, D.C., as Halderman delivered a much-rehearsed line at the onset of a six-minute statement. Until the U-M computer science professor began his testimony before the Senate Select Committee on Intelligence in June 2017, the idea of a hacked American election felt to many lawmakers like a still-theoretical notion. Other technologists and elections integrity experts had warned members of Congress in such formal settings about abstract vulnerabilities, but state officials and election machine vendors had repeatedly insisted they had it all under control.

Halderman has little patience for such coddling. That his voting machine intrusions took place in laboratories rather than live elections made his message no less alarming to the committee.

"We've created attacks that can spread from machine to machine like a computer virus and silently change election outcomes," Halderman continued. "We studied touch screens and optical scan systems." Then, emphasizing each next word with a staccato delivery and direct eye contact,

he stated: "And in every single case, we found ways for attackers to sabotage machines and to steal votes. These capabilities are certainly within reach for America's enemies."

After the aforementioned decade of warning lawmakers about the dangers posed by the machinery of U.S. elections, Halderman, 37, had delivered his message directly to the country's most powerful people. Since then, he has returned to the Capitol routinely to chat with legislators and their staff as Congress passed $380 million in funding for states to modernize their equipment and security practices. In addition, Sen. Richard Burr, the chair of the committee Halderman testified before, sought his input into an election reform package that, as of press time, has yet to be introduced.

It was, he senses, his willingness to declare everything not just hackable but hacked that made heads turn during his Senate testimony. And it is those daring theatrical flourishes—combined with a congenial demeanor of genuine, limitless patience with less tech-savvy people—that has thrust Halderman to the forefront of the quest for safer elections as well as other key high-tech security and privacy issues. At the hearing, Burr, a North Carolina Republican, good-naturedly referred to Halderman as someone who "likes to break in" to election systems and followed up by telling him, "I think what you did was important." Halderman just chuckled along rather than correcting the senator's implication that he'd hacked live elections.

Being likable is one of Halderman's most potent weapons in bending disparate groups of people to his will.

"The archetype sometimes of a technical person might be someone who attaches less significance to the side of cooperating and interacting with people," says David Robinson, Halderman's dorm neighbor during his undergraduate days at Princeton University and now a principal of a Washington, D.C.–based tech policy consultancy. "For Alex, the question of how you align everyone's politics and incentives in such a way that you're going to accomplish something extraordinary together is something that really comes naturally to him."

On a paper napkin, J. Alex Halderman diagrams ways that hackers can infect voting machines with malware. | Photo by Marc McAndrews

In computer science circles, Halderman was a rock star long before he went to Capitol Hill to scare the bejesus out of everybody about the fragility of American democracy. During his first semester as a Princeton graduate student, he and his mentor, professor Ed Felten, showed how easy it was to defeat Sony BMG's efforts to prevent CD piracy.

Not long after, Felten drew the promising young researcher into a project that would go on to inform much of Halderman's career: electronic voting security. After the 2000 election debacle in Florida, with all those hanging chads and confusion about voter intent on paper ballots, Congress gave states more than $3 billion to modernize their voting machinery. This led to a widescale move to touch-screen balloting and computerized tabulations, yet few states or equipment vendors would give independent researchers access to assess how secure these machines were. So in 2006, Felten made contact with an elections insider willing to slip him a commonly used model.

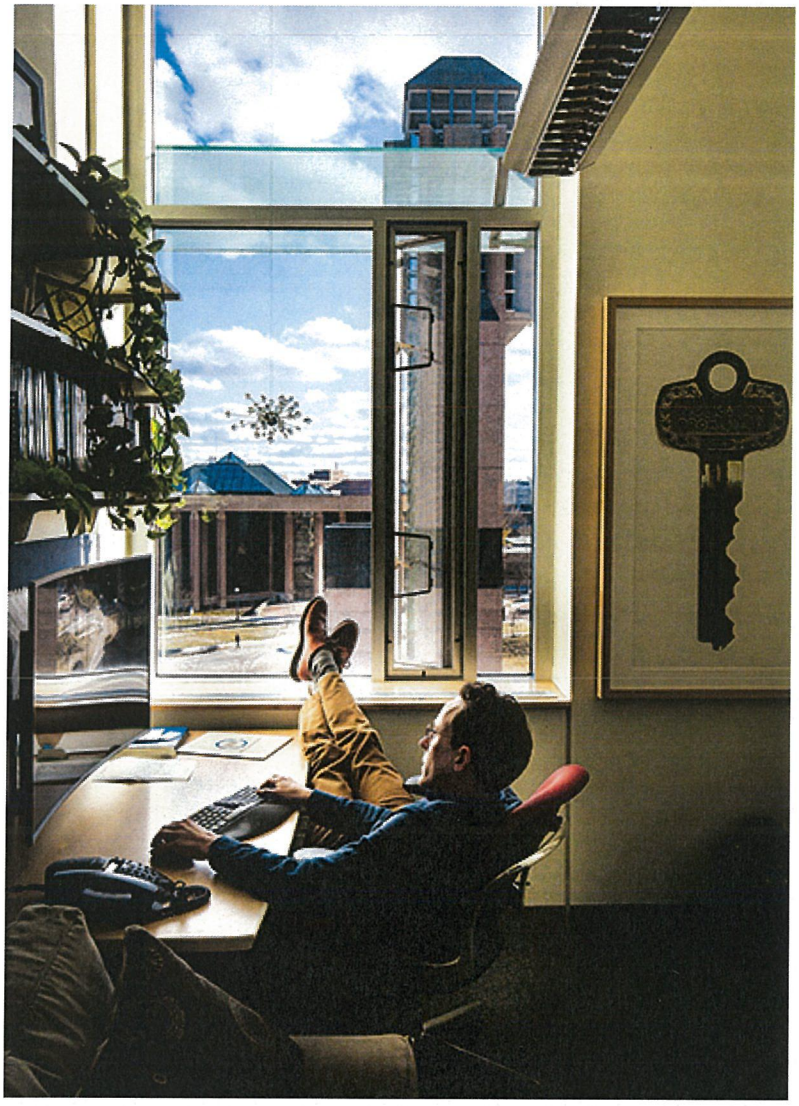
This set up a scene reminiscent of a spy novel, with Halderman, then 25, meeting in an alley with a man in a trench coat who handed him a large leather briefcase containing the contraband voting machine. A few months later, the team posted a YouTube video showing the machine being hacked in a mock election in which Benedict Arnold wins the presidency despite voters clearly choosing George Washington.

That sort of cheeky antic became a signature feature of Halderman's efforts to alert the public to technological insecurities. In 2010, most notably, the District of Columbia was planning to allow citizens to vote via the internet in municipal elections. Online voting is, to Halderman, a particularly terrible idea and one that he has worked against by exposing security flaws in systems used in Australia, Estonia, and Norway.

To demonstrate and test the district's system to the public, the city held a mock election a few weeks before election day. Halderman—in his second year as an assistant professor of computer science at U-M—saw this as "a fantastic opportunity to test out attacks in a live system but not an actual election."

His team easily broke in, altering votes without detection, and even commandeered the video surveillance of the system's servers. In fact, the only reason anyone noticed the breach was the music on the "thank you for voting" page: His students had set the system to play "The Victors."

District officials canceled the online voting idea and never returned to it.



Halderman, director of U-M's Center for Computer Security and Society, in his office | Photo by Joseph Xu/Michigan engineering

One day in 2011, Halderman was at a whiteboard fielding questions from undergraduate engineering students in his "Introduction to Computer Security" class.

A junior asked why a certain approach to circumventing internet censorship in places like China wouldn't work, so Halderman began explaining its flaws. As he did, though, an idea popped into his head. The class, he says, didn't notice the few moments when he stopped and stared at the board, but at that moment a groundbreaking concept now known as "refraction networking" became fixed in his brain. Refraction networking provides a way to deceive censors into thinking they have successfully blocked citizens from banned websites and services while they have, in actuality, allowed access.

Four years after that brainstorm in Ann Arbor, Halderman appeared in New York City with then-United Nations Ambassador Samantha Power to explain the concept at the Internet Freedom Technology Showcase, held alongside the 2015 U.N. General Assembly meeting. Halderman would go on to helm a coalition, relying on more than $2 million in federal funding from the State Department's Bureau of Democracy, Human Rights, and Labor, and this summer the second pilot deployment of the technique took place. Steven Schultze, a former State Department program officer in the bureau, says refraction networking is "a generational jump forward" and "the most promising of all the anti-censorship programs going on."

Halderman, now a tenured professor at U-M and the founding director of the University's Center for Computer Security and Society, describes his eureka moments as instances in which "the pieces snap together. You set up for it and then—aha! When you're working on hard problems, it's not so often when you get beautiful solutions."

Beauty and elegance are traits Halderman clearly treasures, a product of his upbringing on a 50-acre wooded plot in Bucks County, Pennsylvania. His parents indulged his natural itch to disassemble electronics but also took him to New York often to see the opera. He opens some speeches with a portrait of his great-grandfather Maxo Vanka, a prominent Croatian-born artist, and uses that ancestry to trace his own philosophy of promoting security and privacy to Vanka's efforts to fight fascism.

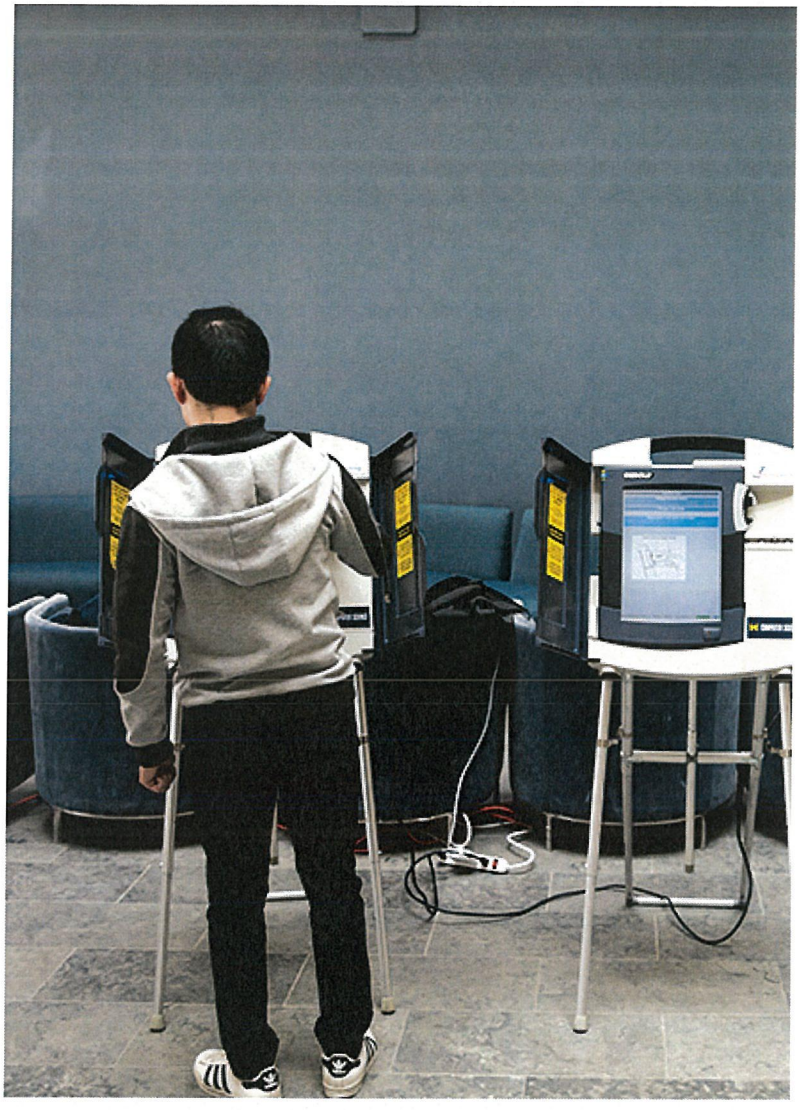
Halderman's office reflects much of his diverse interests and views. His shelves are overwhelmed with works by the likes of Plato and Homer as well as the expected computer science texts and a Geiger counter bought at the Titan Missile Museum in Tucson, Arizona, as a "symbol of a certain era of fear, of where we don't want to go." One telling piece of art on his wall is a poster he made at Princeton showing a blown-up image of a key engraved with the words "DUPLICATION PROHIBITED." "It's the key to the room that contains a giant printer on which it was printed," Halderman says with a smirk. "Using the information in this picture, you can replicate not only the physical key by going and cutting one but the poster of the key by printing one after getting in."

Halderman's long-standing love of the humanities has made him especially aware of the real-world consequences of the misuse of technology. That helps to explain the dramatic array of technological discovery. From his U-M lab, he and his students have alerted Homeland Security that full-body scanners in common use at airports can be effortlessly duped. They also have developed a now-widely used method of querying every IP address in the world in minutes. And they have persuaded the Chinese government to abandon its efforts to require that all computer

users load a piece of surveillance software by demonstrating how vulnerable that made every PC in the country to hack attack.

In 2016, he took a group of students to Hamburg, Germany, for the Chaos Computer Club, billed as the world's biggest hacker conference. There, they watched him and a Princeton colleague reveal to the world that they had figured out the technological approach taken by the National Security Agency to intercept the enormous amount of material it captured according to the documents leaked by NSA whistleblower Edward Snowden. As part of the presentation, Halderman also told the world how best to undermine the NSA's surveillance.

"Alex chooses problems that aren't just academically interesting but have a real-world connection," says Zakir Durumeric, then a U-M doctoral candidate who is now an assistant professor of computer science at Stanford. "If you look at the papers we've written over the last couple years, we're looking at how we can improve security today."



A U-M student votes in the experimental election held on North Campus in a demonstration for The New York Times. | Photo by Levi Hutmacher/Michigan engineering

The only reason there's no evidence of whether voting machines or vote tabulating equipment was hacked in the 2016 presidential election, Halderman insists, is because nobody allowed him or anyone else to check. This is the core of his advocacy regarding electronic voting machines and vote tabulators: He loves technology and believes it can improve lives, but he also urges extra caution when it comes to a process as important as selecting leaders.

In the weeks after the election, Green Party candidate Jill Stein filed for recounts of votes in Michigan, Wisconsin, and Pennsylvania. The intellectual backbone of that effort, however, came from Halderman and a clutch of computer scientists and elections experts who pushed for the chance to analyze the computer equipment used in those states for evidence of malware. After an

erroneous report in New York magazine set off a frenzy by claiming Halderman felt he had "persuasive evidence that the results ... may have been manipulated or hacked," Halderman wrote a widely read Medium essay in which he asserted he never said that but was, nonetheless, concerned.

"The only way to know whether a cyberattack changed the result is to closely examine the available physical evidence—paper ballots and voting equipment in critical states like Wisconsin, Michigan, and Pennsylvania," he wrote. "Unfortunately, nobody is ever going to examine that evidence unless candidates in those states act now, in the next several days, to petition for recounts."

In the end, the effort didn't succeed. On cable news and social media, Halderman was dubbed a Stein puppeteer trying to steal the election for Hillary Clinton, and court rulings blocked recounts in Pennsylvania and halted them in Michigan. In Wisconsin, recounts were completed with negligible vote changes, but nobody was able to inspect any of the equipment.

It was remarkable, then, that just six months later Halderman was invited to testify for the U.S. Senate and received warm reception from members of both political parties in a setting that can be notoriously partisan and contentious. To prepare, Halderman spent a few days with a "murder board" of friends and colleagues drilling him with possible questions and rehearsing his opening statement. The aim was for Halderman to avoid seeming partisan.

"One thing we were careful about was trying to figure out how to keep the focus on the secure voting systems we all want instead of letting the conversation go down a rabbit hole of concern about the election just passed," says Robinson, who helped edit Halderman's testimony. "We didn't want him to mention particular problems. We want everyone to have reason to trust our elections."

On video of the hearing, Halderman appears unflappable as he explains why a certain type of inexpensive, statistically sound audit of paper ballots after an election ought to be routine and is key to double-checking the computer's results. In actuality, he says, "My adrenaline levels were so high, my heart was beating so fast. It was all I could do to read those prepared remarks, but when I was done, it was a tremendous relief."

The message seemed well-received, and a few states are starting to consider post-election audits. Since then, Halderman has become a media fixture. The New York Times even produced a short

film in which Halderman staged a mock election between Ohio State and U-M at the Beyster Building on North Campus. Knowing that most students would vote for U-M, he demonstrated how easy it is to hack the machines and produce a Buckeyes win.

He's still worried about the health of the democratic process, but he tinges his alarm with some optimism. Asked whether the country is any better prepared for the 2018 midterm elections than it was in 2016, he replies, "Oh, it's more or less the same. It's not great news. But, if anything, we're watching more vigilantly. If the systems are probed or attacked, it's more likely we'll find out about it in 2018. Does that mean that attacks won't succeed that would have succeeded before? I don't think we have a basis for strongly increased confidence there. But there are more people watching."

Steve Friess is a Michigan-based freelance journalist and a 2011-12 Knight-Wallace Fellow at U-M. His work appears regularly in The New York Times, The New Republic, Playboy, and many others.

SHARE:

An Artful Space for Science

PREVIOUS NEXT

History Lessons: Parade Days

RELATED POSTS

MAKING TRACKS: LATE FALL 2015

CHANGING LIVES, AND CLOTHING, WITH SLOW FASHION

A NEW INNING AT MICHIGAN AND TRUMBULL

125 YEARS: CELEBRATING MICHIGAN ALUMNUS MAGAZINE

Send Feedback

200 Fletcher St, Ann Arbor, MI 48109 (closed for renovation)

800.847.4764 | alumni@umich.edu

US008033463B2

(12) **United States Patent**
Felten et al.

(10) **Patent No.: US 8,033,463 B2**
(45) **Date of Patent: Oct. 11, 2011**

(54) **SYSTEM AND METHOD FOR MACHINE-ASSISTED ELECTION AUDITING**

(75) Inventors: **Edward W. Felten**, Princeton, NJ (US); **Joseph A. Calendrino**, Ashburn, VA (US); **J. Alex Halderman**, Princeton, NJ (US)

(73) Assignee: **The Trustees of Princeton University**, Princeton, NJ (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1103 days.

(21) Appl. No.: **11/833,955**

(22) Filed: **Aug. 3, 2007**

(65) **Prior Publication Data**

US 2009/0037260 A1 Feb. 5, 2009

Related U.S. Application Data

(60) Provisional application No. 60/952,960, filed on Jul. 31, 2007.

(51) **Int. Cl.**
G06K 17/00 (2006.01)
G06K 13/00 (2006.01)

(52) **U.S. Cl.** **235/386**

(58) **Field of Classification Search** 235/386; 705/12

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2004/0195323	A1	10/2004	Vadura et al.
2005/0247783	A1	11/2005	Poulos et al.
2006/0041516	A1	2/2006	Bogasky et al.
2007/0007341	A1	1/2007	Poulin et al.

OTHER PUBLICATIONS

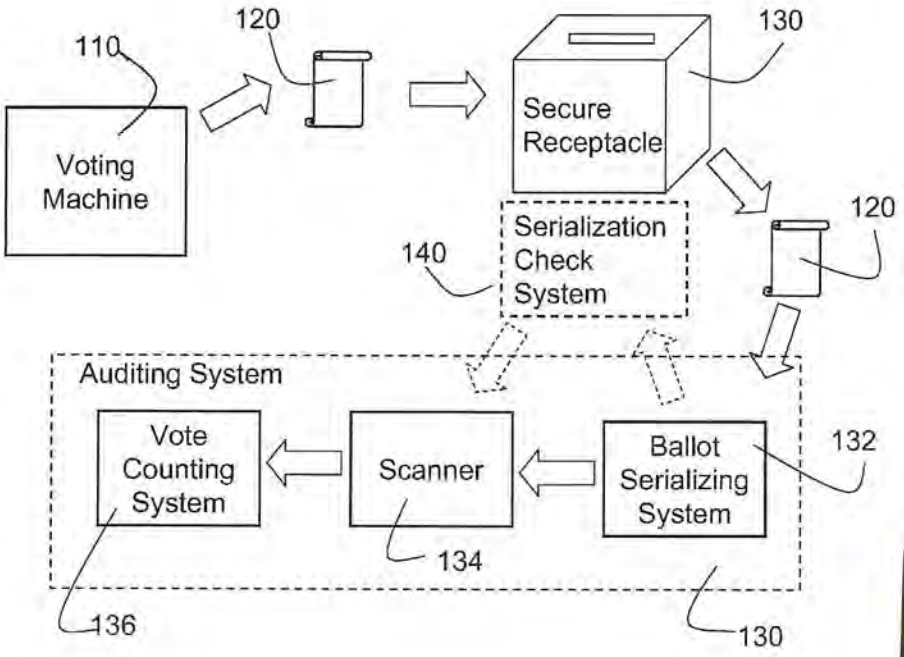
Neff, C. A., "Election confidence: A comparison of methodologies and their relative effectiveness at achieving it," Dec. 2003.
Johnson, K. C., "Election certification by statistical audit of voter verified paper ballots," Oct. 2004.

Primary Examiner — Thien M. Le
Assistant Examiner — Christopher Stanford
(74) *Attorney, Agent, or Firm* — 24IP Law Group; Timothy R. DeWitt

(57) **ABSTRACT**

A method for auditing ballots cast in an election, wherein each ballot is associated with a group. A subset of groups from which sample ballots will be chosen is identified. An identifier is printed on each ballot in the subset. Each ballot has a different identifier than every other ballot in its group. A check is performed to determine whether the identifiers were printed correctly on the ballots. If so, a machine re-count of ballots in each group in the subset is performed and the results are compared to the initial tally of ballots associated with the group. If there is a mismatch, a further investigation is triggered. If there is a match, manual verification is performed on sample ballots from each group. The audit may begin prior to completion of voting from all precincts by estimating the number of samples that will be necessary.

10 Claims, 3 Drawing Sheets



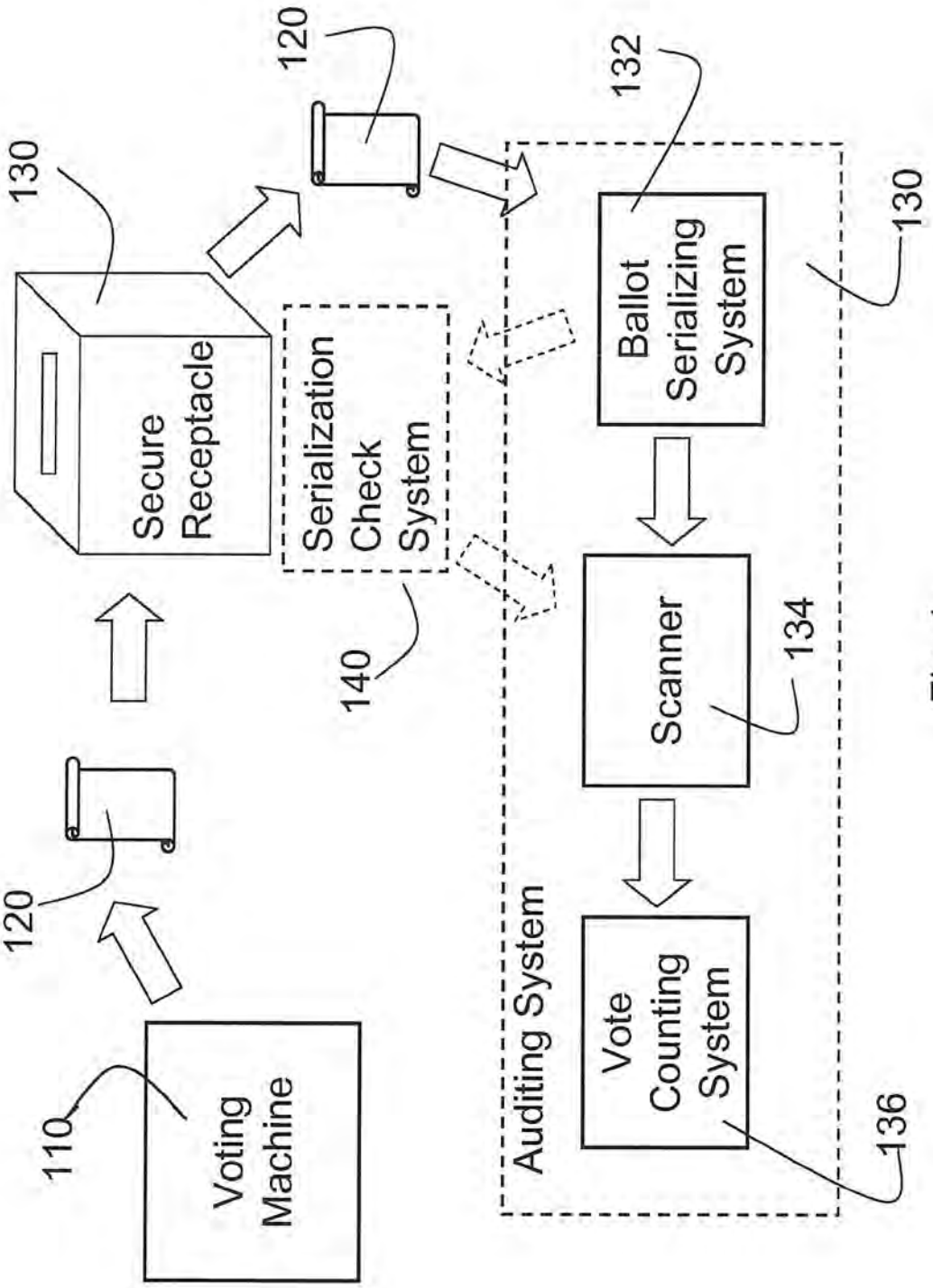

110
Voting Machine
120
120
130
Secure Receptacle
140
Serialization Check System
Auditing System
132
Ballot Serializing System
130
Scanner
134
Vote Counting System
136

Fig. 1

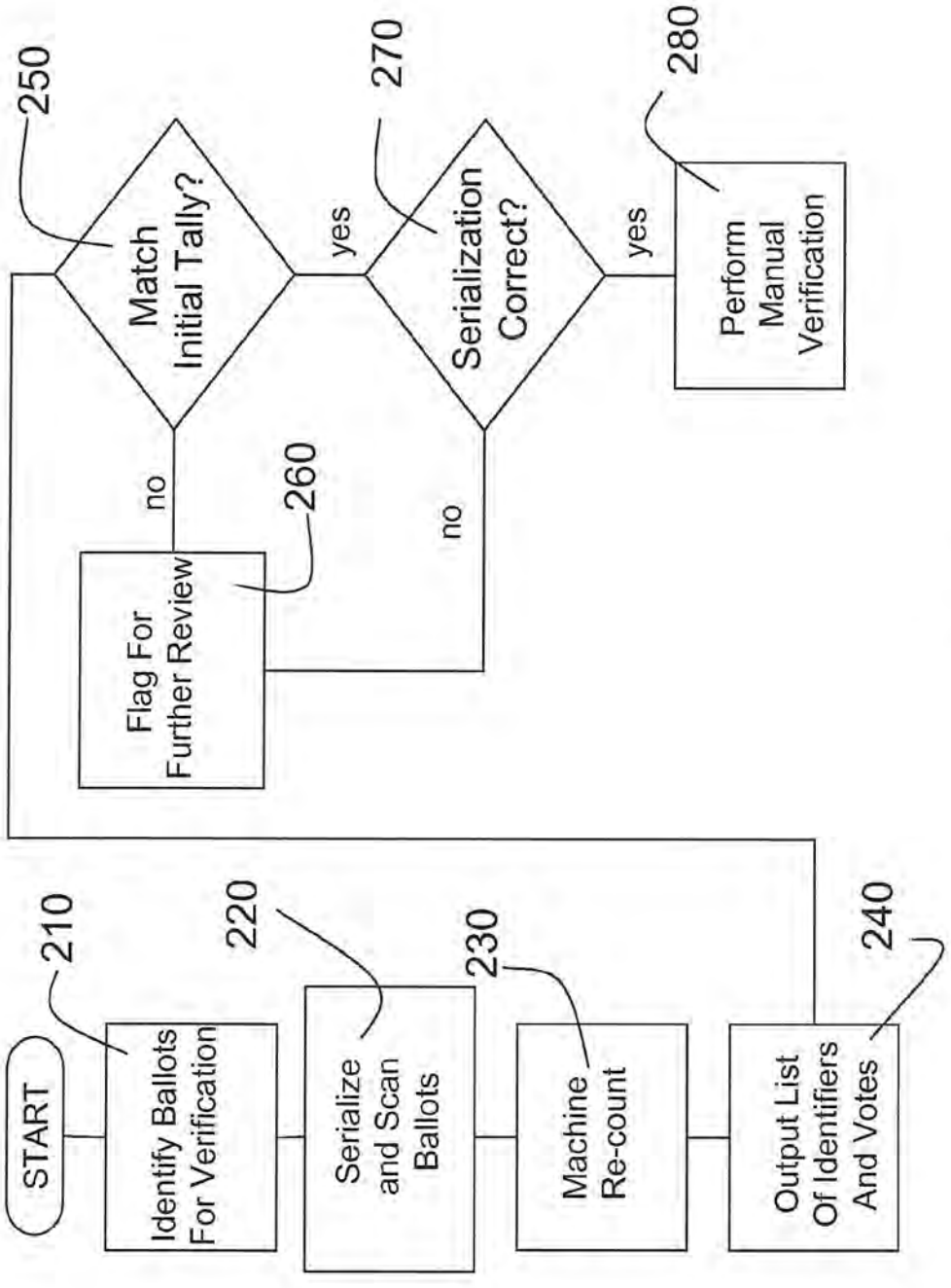

START
210
Identify Ballots For Verification
220
Serialize and Scan Ballots
230
Machine Re-count
240
Output List Of Identifiers And Votes
250
Match Initial Tally?
yes
270
Serialization Correct?
yes
280
Perform Manual Verification
no
260
Flag For Further Review
no

Fig. 2

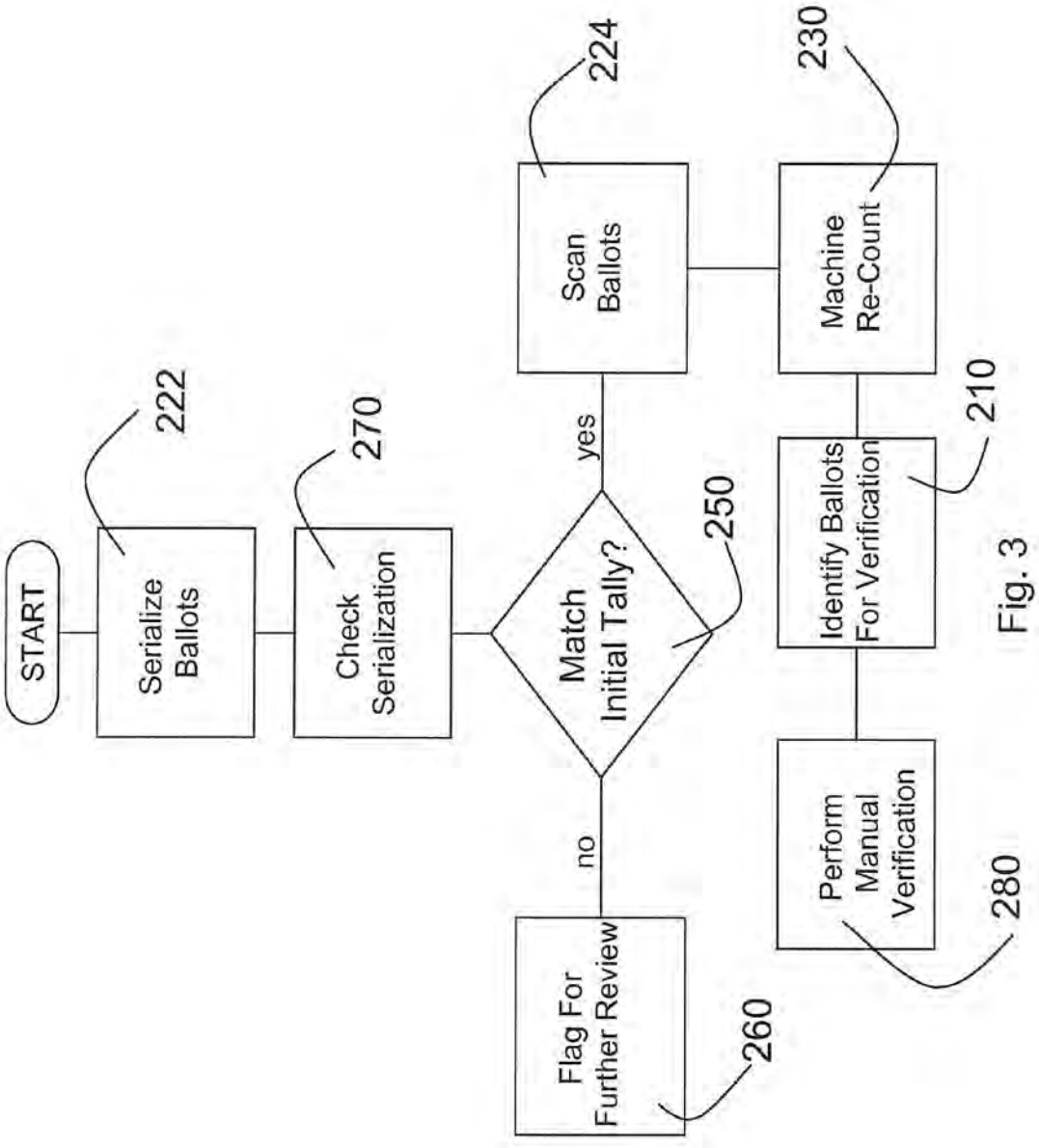

START
Serialize Ballots
222
Check Serialization
270
Match Initial Tally?
250
yes
no
Scan Ballots
224
Flag For Further Review
260
Machine Re-Count
230
Identify Ballots For Verification
210
Perform Manual Verification
280

Fig. 3

SYSTEM AND METHOD FOR MACHINE-ASSISTED ELECTION AUDITING

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of the filing date of U.S. Provisional Patent Application Ser. No. 60/952,960 entitled "System and Method for Machine-Assisted Election Auditing" and filed on Jul. 31, 2007, by Inventors Edward W. Felten, Joseph A. Calandrino, and J. Alex Halderman.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This work was supported by Department of Homeland Security Award No. DE-AC05-060R23100 and National Science Foundation Fellowship 2004016343.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to the field of election auditing.

2. Brief Description of the Related Art

Security analyses of computerized voting systems, including DREs and optical scan machines, have exposed numerous vulnerabilities that could compromise the integrity of elections performed using these devices. See Kohno, T., Stubblefield, A., Rubin, A., and Wallach, D., "Analysis of an electronic voting system," *Proc. 2004 IEEE Symposium on Security and Privacy*, pp. 27-42; Feldman, A., Halderman, J. A., and Felten, E., "Security analysis of the Diebold AccuVoteTS voting machine," *Proc. 2007 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT '07)*. One proposed defense against such attacks is to produce voter-verified paper records and audit them to ensure that they support the totals claimed by the machines.

The most common auditing method is the precinct-based audit, in which workers count all paper ballots from selected precincts and compare the results to the reported precinct tallies. See Appel, A. W., "Effective audit policy for voter verified paper ballots in New Jersey," February 2007; Rivest, R. L., "On estimating the size of a statistical audit," November 2006; Rivest, R. L., "On auditing elections when precincts have different sizes," April 2007; Saltman, R. G., "Effective use of computing technology in vote tallying," Tech. Rep. NBSIR 75687, National Bureau of Standards, March 1975. Unfortunately, performing precinct-based audits can require considerable time, labor, and expense. These costs are multiplied by the complexity of the ballots in many elections, which may include dozens of contests. In a trial recount of a DRE paper trail performed in Cobb County, Ga., workers took an average of 5 minutes per ballot to audit 976 votes at a total cost of nearly $3,000. Dunn, S., "Voter verifiable paper audit trail pilot project," Cobb County, Georgia, November 2006. Unless efficiency can be improved, performing a similar recount of 3% of precincts in New Jersey could cost more than $200,000. Slow, expensive manual recounts limit the level of confidence that can be achieved within a fixed election budget, and they may delay the detection of errors until well after election results have been announced and losing candidates have conceded.

Statistical "ballot-based" audits are an alternative to manually recounting every ballot from selected precincts. Workers sample from all the paper ballots in all precincts and use the sample to assess the accuracy of the original count. Ballot-based audits tend to be more efficient than traditional precinct-based audits, since fewer ballots need to be recounted to achieve the same level of confidence in the result. Neff, C. A., "Election confidence: A comparison of methodologies and their relative effectiveness at achieving it," December 2003. For example, in a statewide race in New Jersey, fewer than one ballot per precinct (4,599 ballots total) would need to be sampled to achieve 99% confidence that the outcome had not been shifted by more than 0.2%. By contrast, over 150,000 ballots (6.9% of precincts) would need to be recounted using standard precinct-based audits (e.g., Stanislevic, H., "Random auditing of voting systems: How much is enough?," August 2006) to achieve the same confidence.

Neff and Johnson were among the first to propose combining ballot-based audit techniques with electronic voting. See Johnson, K. C., "Election certification by statistical audit of voter verified paper ballots," October 2004. Neff assumes that the voting machines link each paper ballot to its electronic counterpart using, for example, a unique identifier printed on the paper ballot and stored with the electronic ballot. When voting is complete, each precinct commits to its set of electronic ballots, then demonstrates that the paper ballots in a given random sample match the corresponding electronic ballots.

The primary weakness of this method is that it establishes the link between electronic and paper ballots at the time that votes are cast. This raises problematic voter privacy issues. For example, if the ballots are linked using sequentially increasing serial numbers, observers could correlate votes with the order in which they were cast, which can reveal the identity of voters. While a cryptographic link might protect privacy, opaque, random-looking identifiers printed on ballots may provide covert channels for leaking voter identities. Even if used securely, they might aid malicious parties who seek to intimidate voters by undermining their confidence in the secrecy of the ballot. Our audit strategy postpones linking paper and electronic records until the recount phase, which allows it to achieve equivalent confidence without jeopardizing privacy or resorting to cryptography.

Johnson alternatively proposes delaying both vote tallying and serial number printing until after all ballots are submitted, allowing voting machines to be simple, memory-less ballot printers. Voters submit their ballots, which, once polls close, are randomized and scanned/tallied. The tallying machine is therefore able to print serial numbers while scanning without privacy risk. Unlike Johnson, we assume that the voting machines maintain an electronic tally, which helps deter traditional attacks against paper-based voting, such as ballot-box stuffing, and, as we will show, provides opportunities for improving the efficiency of the audit.

SUMMARY OF THE INVENTION

The present invention incorporates an alternative audit strategy that substantially reduces these costs by using specialized machines to automate most of the work of recounting paper ballots followed by a manual audit of the machine results. The problem with machines, of course, is that the ones used for the recount are not necessarily more trustworthy than the ones used in the initial count. They may be useful for catching inadvertent errors (especially if they use a different technology and independently developed software), but a determined attacker could still target both sets of machines. What we desire is software independence—an assurance that any tampering with the machines will not cause undetected changes to the election outcome. See Rivest, R. L., and Wack, J. P., "On the notion of 'software independence' in voting

systems," July 2006. To achieve this, we pair recount machines with efficient statistical auditing techniques that allow humans to confirm that the election outcome is correct.

In a preferred embodiment of the present invention a novel audit approach is used wherein ballots are recounted using recounting machines, and their output is manually audited by humans using ballot-based auditing techniques. The efficiency of the method of a preferred embodiment of the present invention is evaluated using data from Virginia's November 2006 elections, and we find that it enjoys significant gains compared to the traditional precinct-based approach. In other embodiments of the present invention, several extensions used to address practical considerations and to further improve efficiency, including means of using knowledge of ballot contents to reduce the sample size.

In a preferred embodiment, the present invention is a method for auditing ballots in an election. Each ballot is associated with a voting machine and with one of a plurality of groups of ballots. The group may be, for example, all ballots from a particular precinct or all ballots from one voting machine. Other groupings of ballots may be used as well. Each group of ballots has an associated initial ballot tally, and each group comprises all ballots associated with at least one voting machine. The method comprises the steps of identifying a subset of groups from which a plurality of sample ballots will be chosen, wherein the subset comprises fewer that all of the groups, printing an identifier on each ballot in each group of ballots in the subset of groups, wherein each ballot in each group has a different identifier than every other ballot in that group, performing a machine re-count of ballots in a group of ballots in the subset, comparing the results of the machine re-count to the initial tally of ballots associated with the group; and flagging the group as containing an error if the machine re-count for the group does not equal the initial tally associated with the group. The steps do not need to be performed in the above-referenced order, and other sequences of the steps will be apparent to those of skill in the art. The method may further comprise the step of determining whether the identifiers were printed correctly on the ballots. Each group may comprise, for example, all ballots associated with one particular voting machine. The method may further comprise the step of identifying a plurality of the groups of ballots from each of which at least one ballot will be selected for manual verification, selecting a plurality of ballots for manual verification and/or performing manual verification of a plurality of ballots. In a preferred embodiment, the identifier is a serial number, but other types of identifiers may be used.

In another embodiment of the invention, the step of identifying a subset of the plurality of the groups is performed before all voting precincts have reported their votes. The step of identifying a subset of the plurality of groups may comprise the steps of estimating a proportion of ballots associated with previously reported groups to a number of total anticipated ballots cast, estimating a minimum number of sample ballots for verification, selecting a plurality of preliminary sample ballots from the previously reported groups wherein the number of preliminary sample ballots selected is greater than or equal to the product of the estimated proportion of ballots and the estimated minimum number of sample ballots, and identifying all of the previously reported groups of ballots having at least one of the preliminary sample ballots.

In other embodiments, the method may further comprise the steps of computing a true minimum number of sample ballots after completion of reporting from all groups, randomly selecting from all ballots a number of sample ballots equal to the true minimum number of sample ballots, comparing how many of the randomly selected sample ballots are associated with the previously reported groups with the number of preliminary sample ballots. If the number of randomly selected sample ballots associated with the previously reported groups is greater than the number of preliminary sample ballots, randomly selecting from the previously reported groups an additional number of sample ballots equaling a difference between the number of randomly selected sample ballots associated with the previously reported groups and the number of preliminary sample ballots. If the number of randomly selected sample ballots associated with the previously reported groups is less than or equal to the number of preliminary sample ballots, performing no further verifications in the previously reported groups.

In another embodiment, the present invention is a method for performing a vote audit comprising the steps of estimating a proportion of ballots associated with previously reported groups to a number of total anticipated ballots cast, estimating a minimum number of sample ballots for verification, selecting a plurality of preliminary sample ballots from the previously reported groups wherein the number of the preliminary sample ballots selected is greater than or equal to the product of the estimated proportion of ballots and the estimated minimum number of sample ballots, identifying all of the previously reported groups of ballots having at least one of the preliminary sample ballots, performing vote verification in each identified previously reported group, computing a true minimum number of sample ballots after completion of reporting from all groups, randomly selecting from all ballots a number of sample ballots equal to the true minimum number of sample ballots, and comparing how many of the randomly selected sample ballots are associated with the previously reported groups with the number of preliminary sample ballots. If the number of randomly selected sample ballots associated with the previously reported groups is greater than the number of preliminary sample ballots, randomly selecting from the previously reported groups an additional number of sample ballots equaling a difference between the number of randomly selected sample ballots associated with the previously reported groups and the number of preliminary sample ballots. If the number of randomly selected sample ballots associated with the previously reported groups is less than or equal to the number of preliminary sample ballots, performing no further verifications in the previously reported groups. The step of performing a vote verification in each identified previously reported group comprising the steps of performing a machine re-count of ballots associated with each identified previously reported group and performing manual verification of each preliminary sample ballot associated with each identified previously reported group. The step of randomly selecting ballots comprising randomly selecting each ballot with equal probability or randomly selecting ballots with weighted probabilities. In another embodiment, the step of estimating a minimum number of sample ballots for verification may based upon an expected number of switched ballots necessary to change the outcome of an election.

In still another embodiment, the present invention is a method for performing a vote audit of ballots in an election comprising the steps of calculating a minimum number of sample ballots for verification based upon the number of the ballots and the contents of the ballots, selecting sample ballots for verification based upon the contents of the ballots, performing verifications of the sample ballots.

In another embodiment, the present invention is a system for performing a vote audit that comprises means for identifying a subset of groups from which a plurality of sample ballots will be chosen, wherein the subset comprises fewer

than all of the groups, means for printing an identifier on each ballot in each group of ballots in the subset of groups, wherein each ballot in each group has a different identifier than every other ballot in that group, mean for performing a machine re-count of ballots in a group of ballots in the subset, means for comparing the results of the machine re-count to the initial tally of ballots associated with the group; and means for flagging the group as containing an error if the machine re-count for the group does not equal the initial tally associated with the group.

Still other aspects, features, and advantages of the present invention are readily apparent from the following detailed description, simply by illustrating a preferable embodiments and implementations. The present invention is also capable of other and different embodiments and its several details can be modified in various obvious respects, all without departing from the spirit and scope of the present invention. Accordingly, the drawings and descriptions are to be regarded as illustrative in nature, and not as restrictive. Additional objects and advantages of the invention will be set forth in part in the description which follows and in part will be obvious from the description, or may be learned by practice of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention and the advantages thereof, reference is now made to the following description and the accompanying drawings, in which:

FIG. **1** is a block diagram of a vote auditing system in accordance with a preferred embodiment of the present invention.

FIG. **2** is a flow diagram of an overview of a preferred embodiment of a method in accordance with the present invention.

FIG. **3** is a flow diagram of an overview of an alternate preferred embodiment of a method in accordance with the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In a preferred embodiment of the present invention, manual precinct-based audits are replaced with machine-assisted audits. As shown in FIG. **1**, voters cast votes at a voting machine **110**. The voting machine may be of any type, electronic or otherwise. When a vote is cast, a hard copy record **120** of the vote is generated, by hand or by the voting machine. For privacy reasons, the hard copy record preferably does not have any information, or has as little information as possible, that could be used to associate that ballot with a particular voter. The hard copy record **120** is placed into a secure receptacle **130**. When a re-count or audit of the votes is to be performed, the hard copy records are fed into a re-count or auditing machine **130**. The vote auditing machine may have, for example, a ballot serializing system **132** for assigning and printing identifiers on the hard copy records, a scanner **134** for scanning the hard copy records, and a vote counting system for counting the votes on the scanned ballots. As discussed below, a serialization or identifier check preferably is performed manually, but alternate embodiments with an electronic checking system **140** also are possible with the present invention. When a recount of a group of votes is to be performed, poll workers, rather than recounting ballots manually, feed them through a specialized recount or auditing machine **130** that functions like a combined optical scanner **134**, printer or ballot serializing system **132**, and vote counting system **136**.

In a preferred embodiment of the method of the present invention will be described with reference to FIG. **2**. After a group of ballots is identified for a re-count at step **210**, the auditing machine **130** scans the contents of each ballot and prints an identifier on the each ballot **220** and stores the identifier along with the associated ballot contents. The identifier may be a unique serial number or may be any other identifier that permits each ballot to be distinguished from other ballots in a similar group. For example, each ballot from one particular precinct could receive and identifier that distinguishes it from all other ballots from that particular precinct. In this example, the identifiers do not need to be different than identifiers assigned to ballots from other precincts. Similarly, each ballot from one voting machine could contain an identifier that distinguishes the ballot from all other ballots from that particular voting machine. In other words, the identifier may be truly unique or may just be unique within a particular group of ballots, such as all ballots from one precinct or all ballots from one voting machine. The present invention may be used with other groupings of ballots and such other groupings will be apparent to those of skill in the art.

At the end of the scanning process, the machine re-counts the votes **230** and outputs a list of votes on each ballot together with the ballot's identifier **240**. While this is referred to in FIG. **2** as serializing, it should be understood that identifiers other than serial numbers may be used. The results of the re-count are compared to the initial tally of votes **250**. If the re-count tallies differ from the initially reported electronic count, discrepancies clearly exist and a wider investigation should be conducted **260**. Depending on circumstances, an appropriate response might be to inspect the corresponding machines, other machines of the same model, other ballots in that precinct, etc. Other responses will be apparent those of skill in the art.

If both tallies match, the workers perform a secondary audit **270** to check the accuracy of the machine's recount. For example, with sequential serial numbers they may first quickly flip through the pile of numbered ballots to ensure that it increases sequentially from one to the reported ballot total without repeats. This check helps protect against collusion between voting and recount machines, as described shortly. Other means for ensuring that the number of paper and electronic ballots match may be used, such as an electronic check of the identifiers. If the check reveals some type of inaccuracy, the group of ballots are flagged for further investigation **260**. If the check confirms the identifiers have been printed (and/or assigned) accurately, a manual verification is performed **280**. For example, poll workers may take a random sample of the electronic ballot records, retrieve the corresponding paper ballots, and verify that they match.

In a preferred embodiment in which the identifiers are serial numbers, since the ballots are serialized and fed out of the machine in order, retrieving a particular ballot for verification requires very little effort. The most significant labor required may be to check for repeats, which given sequential ordering, is a rapid single-pass process.

In practice, separate devices may be used to perform the printing and scanning functions of the recount machine. When voting is complete, a printer device could place serial numbers on the ballots, and then a separate scanner could read the numbers along with the votes. In precincts utilizing optical scan machines, properly designed machines could perform both the initial count and the recount: this option

decreases costs but reduces redundancy. If the same machine performs counts, recounts, and printing, officials must have some means of mechanically disabling the printer while polls are open, such as removal of the printer head. Printers also must be physically unable to alter the record of the vote on the ballot. They could be designed so that they cannot reach outside of a predefined empty margin on ballots, or they could utilize a kind of ink that would be immediately apparent when ballots were inspected.

Further, various steps in the method may be performed in other sequences or may be split into multiple steps, such as serialization **222** and scanning **224** of ballots, as shown in FIG. **3**. Other variations in the sequence of steps will be apparent to those of ordinary skill in the art.

Security

The redundancy of combining electronic and paper-based systems increases the security of the overall system. With high probability, the manual audit process detects any discrepancies between the sets of electronic and paper ballots that are substantial enough to impact the election's outcome. Because the process checks the correspondence between the sets of ballots, measures improving the integrity of either set increase the overall integrity of the election result. Since both the electronic and paper ballot sets must remain similar for a discrepancy to avoid detection, combined systems are more likely to detect malfunctions, and they increase the sophistication necessary to commit fraud.

For an error to go undetected, the voting machine must report an incorrect electronic tally, the recount machine must support the incorrect tally, and the manual audit process must not detect a discrepancy between the paper and electronic ballot records.

A malfunctioning or dishonest voting machine may add, subtract, or switch votes to introduce errors in its electronic tallies. If election officials maintain an accurate sign-in list for the precinct, any significant discrepancy in the total number of reported ballots and, consequently, votes will be detected. Therefore, the voting machine is limited to switching votes from one candidate to another.

For a recount machine to support an incorrect electronic tally, either the set of paper ballots must match the incorrect tally, or the machine must fail to detect a discrepancy. The set of paper ballots can only match the tally if either the voting machine printed an incorrect set of paper ballots or another party modified that set. If voters generally verify their paper ballots, the ballot box will likely contain an accurate paper ballot for most voters when polls close. While the voting machine may print additional, incorrect ballots, this would cause the number of paper ballots to exceed the electronic ballot total, which reflects the number of voters, so an accurate recount machine would detect this discrepancy. The simple, sequential nature of machine-assisted auditing also reduces opportunities for adversaries to modify paper ballots during the audit. Assuming that no adversary can modify the set of paper ballots, only recount machine malfunction, whether accidental or malicious, would allow the discrepancy to go undetected.

A malfunctioning recount machine may report incorrect electronic ballots that agree with any incorrect electronic tally regardless of the true paper ballots. The machine may even collude with other parties by omitting or printing incorrect serial numbers on paper ballots to hide errors. For example, a voting machine may print additional paper ballots with desirable votes, and a recounting machine may reuse serial numbers on certain undesirable voter-verified paper ballots to effectively replace them with the additional ballots. The manual check of serial numbers detects duplicate or omitted serial numbers and ensures that the number of paper ballots matches the total reported number of electronic ballots.

If no errors are detected before the sampling phase, we know that we have a set of electronic ballots from the recount machine that supports the initial electronic tally and an equal-sized set of paper ballots with corresponding serial numbers. We designed the sampling process specifically to detect discrepancies between these sets significant enough to affect the election's outcome. Unless an error or adversary modified both the initial electronic tally and the paper ballots, the manual audit should catch any remaining errors with a high level of confidence.

Privacy

The present invention avoids many of the privacy issues inherent in some earlier ballot-based audit methods that involve placing identifiers on ballots during the voting process. In the present invention, the ballots do not receive serial numbers or other identifiers until the recount phase, so they are likely to become at least partially reordered before being numbered. Well-designed ballot boxes and cut-and-drop paper trail systems assure that the papers are somewhat shuffled as they are inserted. Since voters widely trust these methods to frustrate correlation with voter check-in times, this provides significant practical privacy benefits. Should alternative ballot shuffling methods offer greater protection, officials may substitute such methods without modifying the audit process. In any case, the recount machine has no more information about the order of votes than would workers performing a manual recount.

Another benefit of this technique is that a voting machine need only maintain tallies rather than electronic copies of individual ballots. These tallies preferably include the total number of ballots submitted and the total number of votes for each option. Thus, voting machine designers do not need to worry about properly shuffling electronic ballots to protect voter privacy or about maintaining storage for those ballots. However, if the same machines perform counts and recounts, which may be the case with the present invention, they must have some means of attaching extra memory during the recount for storing the ballot scan results.

What to Audit

Due to the popularity of plurality voting systems in the U.S. we consider those systems in the preferred embodiments, though machine-assisted audits may be useful in many other voting systems. With plurality voting, voters may choose a number of candidates equal to the number of seats available. (This is a mild misuse of the term plurality system: other forms of plurality voting for multiple candidates exist. See Ace Electoral Knowledge Network, Plurality/majority systems, 2006. http://aceproject.org/aceen/topics/es/esd/esd01/ .) If k seats are available, voters may select up to k candidates, and candidates receiving the top k vote totals are the victors. This definition is an extension of the familiar single-seat contest.

An audit process need only sample enough ballots to confidently detect the minimum amount of fraud that would have affected the election's outcome. To modify the fewest ballots while changing the outcome, an adversary would swap the positions of the losing candidate with the most votes and the victor with fewest votes. Switching votes directly between these candidates requires the fewest ballot changes, as each switch alters the relative difference by two. To do so, the adversary would take ballots with votes for candidate A but not B and change them to contain votes for B but not A. Therefore, we need only audit enough ballots to discover

fraud that alters a number of ballots equal to half the difference (rounded up) in vote totals between the "just losing" and "just winning" candidates.

Two techniques are described for selecting which ballots and precincts need to be audited. The first technique has the benefit of a constant sample size given the number of ballots, the margin of victory, and the desired level of confidence. Sample size may vary with the second approach, but that approach is more amenable to extensions of the present invention discussed below.

Constant Sample Size Method

The hypergeometric distribution describes the number of bad ballots an auditor can expect to find when sampling without replacement. Assume that auditors desire a confidence level c that no fraud significant enough to change the election's outcome occurred. Given N total ballots and a minimum of B incorrect ballots, the probability mass function of the hypergeometric distribution dictates a minimum sample size, n, of:

$$n = \min\left\{u \,\middle|\, 1 - \prod_{k=0}^{u-1} \frac{N-B-k}{N-k} \geq c\right\} \quad (1)$$

A simple computer program can rapidly, verifiably calculate n for any practical value of N.

After all precincts report their recount results and scanned ballots, state officials randomly select n ballots to check. To do so, officials assign each ballot an equivalent portion of the range of a pseudorandom function. Since sampling occurs without replacement, the officials must alter the assignments appropriately after each draw. Representatives for all candidates or issues in a race may assist in randomly generating a key for the function (for example, consider Cordero, A., Wagner, D., And Dill, D., "The role of dice in election audits extended abstract," *IAVoSS Workshop on Trustworthy Elections* 2006). The state then evaluates the function, with the randomly generated key, for the numbers one through n.

Because officials select ballots at random with respect to any given race, officials may use the same ballot from auditing one race in auditing any other race appearing on that ballot, provided that all voters eligible to vote in the latter race are also eligible to vote in the former. This reduces the number of ballots to retrieve. Note that the correlation between votes on a given ballot prevent us from gaining additional assurance from using the same ballots for multiple races, but officials still gain confidence c in the results of each race.

A machine recount of a precinct is only necessary if a ballot will be selected for manual verification in that precinct. Thus, auditors could use the initial electronic tallies to perform a mock ballot selection before the machine recount. Any precinct which would have contained a chosen ballot given the mock selection will undergo a machine recount. Following the machine recount, representatives must generate a new key. Because a pseudorandom function is deterministic, an adversary with knowledge of the key prior to the machine recount could determine which serial numbers will be sampled for manual review following the recount. Such an adversary could collude with the recount machine to hide fraud under serial numbers that will not be sampled Officials may then randomly select a single ballot from each recounted precinct and randomly draw the remaining required ballots from the full pool in all recounted precincts.

Varying Sample Size Method

Rivest proposes an efficient precinct-based auditing technique in which, rather than drawing a given-size sample from the population of precincts, auditors instead randomly select each precinct with a given probability. The same idea is also useful in the context of ballot-based auditing. Assume that, to change the results of an election, the set of ballots must contain a minimum of B bad ballots. To achieve a confidence level of c that at least one bad ballot will be sampled, auditors may select each ballot with probability p chosen such that $(1-p)^B \leqq 1-c$, or $p \geqq 1-(1-c)^{1/B}$.

Officials may follow the same process as before for generating a key and may apply a pseudorandom function to a unique identifier for each ballot (for example, 1 to N, where N is the total number of ballots in all precincts voting on the given issue), mapping the result back to [0,1] to determine whether to check the ballot.

To determine which precincts need to be audited, we may calculate the probability that one or more of the v_i ballots in precinct i will be sampled as $1-(1-p)^{v_i}$. Auditors may select each precinct based on the probability that it contains a sampled ballot. If so, officials perform a machine recount in that precinct. Given that at least one ballot is sampled in a precinct, the probability of sampling k ballots in that precinct is:

$$\frac{\left(\frac{v_i}{k}\right)p^k(1-p)^{v_i-k}}{1-(1-p)^{v_i}} \quad (2)$$

Following the machine recount, officials randomly select the precinct's sample size based on this distribution. As before, officials should generate a new key immediately following the machine recount of selected precincts.

Comparison to the Method of Rivest

Assume use of the audit method in Section 3.2, and let $p=1-(1-c)^{1/B}$. The probability that precinct i requires a machine recount is therefore $p=1-(1-c)^{v_i 1/B}$. If an adversary can steal any number of votes in a precinct without generating suspicion, Rivest proposes a logistic precinct-based approach that yields the same precinct audit probability. For machine-assisted auditing, however, auditors need only manually review a subset of the recounted ballots.

Rivest presents his logistic approach as a non-optimal heuristic, so the usefulness of this link seems limited. Furthermore, the percentage of votes in a precinct that one may steal without generating suspicion is more likely 10-20% than the 100% assumed here. In light of this, a performance comparison between Rivest's optimal precinct-based techniques and our methods under realistic circumstances would be informative.

Evaluation

To evaluate the efficiency of machine-assisted auditing (and ballot-based auditing in general) versus precinct-based auditing, we consider both techniques in the context of available data from Virginia's November 2006 elections, both for local and statewide races. In this example, we considered all races from the available Virginia data. Some local races are absent, so we ignore those. Due to minor absences in the data set, we assume that no voter submitting a ballot abstains from voting on an issue and that voters for multi-seat races submit multiple ballots rather than a single ballot with multiple selections. While these assumptions slightly affect the realism of the tests, they likely had only a minor impact on the overwhelming results.

In all cases, we seek a 99% confidence level. For machine-assisted auditing, we use the techniques discussed above. For precinct-based auditing, we use the methods and assumptions in: auditors choose precincts uniformly at random, an adversary may switch no more than a set percentage of the votes in a precinct without arousing suspicion (we use 10%), and the adversary may switch votes in the largest possible precincts

Virginia contains 2,599 precincts and approximately 4.6 million registered voters, nearly 53% of whom cast ballots during the November 2006 election. The general election decided nineteen issues: four statewide issues, including a U.S. Senate race and several statewide initiatives, and fifteen smaller races, such as U.S. House races. In addition, voters considered numerous local ballot issues, ranging from city council elections to school constructions projects. Virginia State Board Of Elections, General election—Nov. 7, 2006. http://www2.sbe.virginia.gov/web docs/Election/results/2006/Nov/htm/index.htm. Because auditing is typically both more important and more labor-intensive in closer races, we focus on such races, excluding consideration of races for which modification of 10% or more of the ballots would have been necessary to change the outcome. This choice rules out many of the races but leaves a set of 49 remaining. Seven of those remaining were general election issues and forty-two were local issues.

The remaining general election issues include a U.S. Senate race with a margin of victory of 0.39%, four U.S. House races, a race for the Virginia House of Delegates, and a state constitutional amendment. For those races, machine-assisted auditing would require a manual review of approximately 437 ballots on average—0.06% of the 796,469 average total ballots. Only the smaller House of Delegates race would require review of greater than 1% of the ballots (1.05%), and five of seven races require audit rates under 0.1%. Precinct-based auditing would review approximately 177,849 ballots on average—22.33% of the average total ballots. In each case, precinct-based auditing requires an expected hand count of more than 42 times as many ballots. The closely contested U.S. Senate race would require review of 2,337 of 2,370,445 ballots with machine-assisted auditing and 1,141,900 ballots on average with precinct-based auditing.

While less overwhelming, the results for local ballot issues are highly favorable as well. In this case, machine-assisted audits would review approximately 224 ballots on average—2.28% of the 9,842 average total ballots. Precinct-based audits would require manual review of approximately 3,928 ballots on average—39.91% of the average total ballots. Only five of the forty-two races would require a manual review of more than 50% of the ballots with machine-assisted audits. In contrast, only six of the forty-two races would require a review of less than 50% of the ballots on average with precinct-based audits. Precinct-based audits would require a complete recount in more than half of the cases.

The races that are particularly difficult for machine-assisted auditing are town council, city council, and school board races with 7/492, 5/849, 12/769, 7/246, and 3/2409 margins of victory—requiring manual review of 68.3%, 78.4%, 53.4%, 68.3%, and 90.0% of ballots respectively. In each of these cases, precinct-based auditing would require a full recount.

If comparing machine-assisted audits and precinct-based audits purely on the number of manual ballot reviews, these results indicate a conclusive advantage for machine-assisted audits.

Extensions

In this section, we consider a number of methods for increasing the efficiency, practicality, and utility of machine-assisted audits.

Handling Misreadings

With some small probability, auditors might misread a paper ballot and falsely conclude that it either does or does not match the corresponding electronic ballot. Accidentally concluding that the two versions of a ballot do not match is not an issue: auditors would certainly immediately double-check any such ballots. The opposite error would be more serious. We would expect its probability to be low, however, especially in larger elections. In that case, the number of ballots to check per precinct will often be relatively small, meaning that auditors are less likely to become careless. In addition, the state may request and double-check copies of the paper ballots against the reported electronic ballots.

If auditor error is a serious risk, Johnson offers a starting point for adapting sample sizes to overcome such errors, assuming use of the ballot-based audit techniques discussed above. If officials are instead using the precinct-based methods discussed above, these errors are easy to manage. Suppose that an auditor misclassifies a mismatch as a match with probability m. In this case, the true probability of detecting a bad ballot will not be p but will instead be $p(1-m)$. Thus, p must be chosen such that $p \geqq [1-(1-c)^{1/B}]/(1-m)$.

Early Returns

A variety of circumstances may result in delayed reporting from certain precincts. Precincts that report in a timely manner might wish to begin the audit process without waiting hours or days for a complete initial tally. Given partial returns, auditors may assume reasonable or worst case scenarios from the remaining precincts and begin the audit under those assumptions. Once all precincts have reported, unexpected results might force additional sampling from previously reported precincts, but the bulk of the audit process may already be complete.

Assuming the ballot-based methods discussed above, this means that auditors could estimate both the proportion of ballots cast in previously reported precincts (q) and the necessary sample size (n'). They may then select $d \geqq n'q$ ballots from the already-reported precincts. After reporting is complete, auditors could compute the true sample size, create a one-to-one mapping between all reported ballots and $\{1, \ldots, N\}$ and randomly select n values from that set. If d or fewer of the selected values correspond to ballots in previously reported precincts, no additional sampling is necessary in those precincts. If more than d values correspond to those precincts, that number minus d additional ballots must be drawn from the precincts. Similarly, auditors must select a number of ballots from the late reporters equal to the number of selected values corresponding to ballots in those precincts.

Using the precinct-based methods discussed above, auditors would calculate p' based on the expected number of switched ballots required to change the outcome and begin sampling. Once all results are reported, officials may calculate the true value of p and use it for newly reported ballots. If the final margin of victory is smaller than expected, they also must sample previously reported but unsampled ballots with probability

$$p'' = 1 - \frac{1-p}{1-p'},$$

yielding an over/all selection probability of p for those ballots.

Varying Probability by Precinct

In a precinct-based method, the sampling process selects each ballot with equal probability. That process need not do so. For example, officials may prefer to reduce the probability that ballots are selected in extremely small precincts, thereby reducing the probability that machine recounts (and recount machines) will be necessary for a large number of small precincts. The only constraint that the audit process must satisfy is that, given any set of ballots of size B or larger with corresponding selection probabilities $p_1, \ldots, p_B$, $(1-p_1) \ldots (1-p_B) \leqq 1-c$. Thus, auditors may entirely ignore small precincts in some cases provided that they increase the probability of sampling ballots in other precincts to compensate.

Hybrid Strategies

The line between precinct-based auditing and machine-assisted ballot-based auditing need not be so fine. States could use machines to perform recounts in randomly selected precincts then audit the machine results. The embodiment of the previous section technically does this, but the probability of selecting a precinct need not be directly based on the probability of selecting its underlying ballots. This is similar to a dial that auditors could turn. Assume a desired confidence level c. At one extreme, machines could recount all precincts, and auditors could sample ballots such that the overall probability of uncovering a bad ballot—if enough are bad to impact the outcome—is c. This is a machine-assisted ballot-based audit approach.

At the other extreme, auditors could select precincts such that the probability that at least one selected precinct contains a bad ballot—if enough are bad to impact the outcome—is c, and auditors could manually check all ballots in those precincts. This is precinct-based auditing. In either case, the probability of detecting fraud significant enough to affect the election's outcome is c. Between these two possibilities, one could trade a greater expected number of machine recounts for a smaller expected proportion of manual recounts and vice versa to achieve a confidence level of c. Depending on the costs and benefits of each, states may choose whatever balance is most appropriate for their specific circumstances.

Considering Ballot Contents

Consider a two-candidate mayoral race in which the electronic results indicate that Alice beat Bob 10,001 to 10,000. Traditional audit techniques would require that officials consider ballots containing votes for either candidate even though the primary objective is to discover whether any votes for Alice should have been for Bob. Examining only ballots reported to contain votes for Alice could cut auditor work nearly in half, as auditors seek to discover an equivalent amount of fraud in a far smaller pool of ballots. In general, by considering the contents of ballots, officials may reduce the number of manual verifications required. Note that, to apply these methods, either a full machine recount of all precincts or some other means of obtaining a full set of electronic ballots is necessary. For the remainder of this section, we assume use of the above described precinct-based audit process.

Assume a race in which n candidates are competing for k seats, and let $v_1, \ldots, v_n$ be the vote totals for the candidates in decreasing order. Therefore, $v_1, \ldots, v_k$ correspond to winning candidates. Because a single ballot may contain votes for up to k candidates, we need to consider the combination of votes on each ballot. Let C_s, where $1 \leqq s \leqq k$, be the winning candidate with the lowest vote total that received a vote on the ballot. Let C_t, where $k+1 \leqq t \leqq n$, be the losing candidate with the highest vote total that did not receive a vote on the ballot. We need to look for several possibilities:

If C_s exists, we must consider the possibility that the ballot contains a fraudulently added vote for any winning candidate receiving a vote on that ballot. At least $v_s - V_{k+1}$ votes would need to be added to move such a candidate from a losing position to a winning position.

If C_t exists, we must consider the possibility that the ballot contains a fraudulently removed vote for any losing candidate without a vote on that ballot. At least $v_k - v_s$ votes would need to be removed to move such a candidate from a winning position to a losing position.

If C_s and C_t both exist, we must consider both previous cases along with the possibility that a vote was fraudulently switched from a losing candidate without a vote on that ballot to a winning candidate receiving a vote. At least

$$\left|\frac{v_s - v_t}{2}\right|$$

votes would need to be switched to swap the relative positions of two such candidates.

If neither C_s nor C_t exist, the ballot could not have been part of fraudulent activity that changed the election's result, so we may ignore it. Otherwise, let B equal the minimum amount of fraud necessary for any of the applicable cases above. As before, the ballot must be manually checked with probability p such that $p \geqq 1-(1-c)^{1/B}$.

If ballots contain votes for more than one race, we must manually check ballots with the maximum probability necessary for any individual race/vote combination on that ballot.

Considering Initial Returns

Similar tricks may also be useful given only reported initial electronic vote tallies. A precinct in which initial tallies indicate that all ballots contain votes for Bob could not have contributed to discrepancies affecting the election's outcome, so both machine-assisted and precinct-based auditing could ignore that precinct entirely. In a single-seat race, a ballot may contain a single vote at most, so we may determine the precise ballot contents for that race from the initial electronic tallies alone and use that to calculate an appropriate probability of manually verifying each ballot.

As we increase the number of available seats, competing candidates, and races on a ballot, inferences tend to become more difficult and less beneficial. The added complexity is a result of an increase in the number of possible vote combinations on a ballot. One may still draw inferences from the ballots, however. For example, if 43% of ballots contain a vote for the "just losing" candidate, we know that 43% of the ballots could not have had a vote for that candidate removed or switched to another candidate. As vote totals in complex multi-seat races become more tightly clustered, the complication of drawing inferences might counterbalance the increasingly minor benefits of those inferences. A test of these methods on real elections data might help to better establish which cases benefit from these techniques.

Write-Ins

Machine-assisted audits can easily handle the case in which all write-in candidates are put in a machine-readable form, whether by voters or by election officials. They also may handle the case in which voters may not write-in a candidate already appearing on the ballot by treating all write-in votes as votes for a single additional candidate. Otherwise, if the number of write-in votes is insufficient to affect the outcome of the election (given initial tallies), we may assume whatever combination of write-in votes results in the closest possible election and search for the necessary quantity of fraud in ballots not containing write-in votes.

If write-in votes could change the results of the election, a count of those votes will ultimately be necessary before certification of the election. Given that the count of ballots with write-ins is manually performed, we may simply audit the remaining, machine-recounted ballots to discover any fraud large enough to affect the outcome. If write-in ballots are machine-scanned, we may add a serial number printer to that scanner and sample ballots from the full pool. Given the relatively small quantity of write-ins in many elections, we expect this to rarely be a significant issue in practice.

Machine Malfunction

Presumably, some percentage of recount machines will occasionally fail. While a failure would cause a delay in the audit process for the affected precinct, the delay would only be for that precinct. All other precincts could proceed normally, and the affected precinct could wait for repairs or obtain a recount machine from a completed precinct. If voting machines fail, the failures could delay initial tally reports from the affected precincts. As discussed above, such a delay need not hold up other precincts.

Candidate Assurance

To give candidates additional assurance that the audit process did not miss or under-sample precincts in which fraud seems apparent, Appel introduces the idea of allowing candidates to select a small number of additional precincts and pay for full manual recounts of those precincts (candidates are reimbursed if errors are uncovered). We would recommend that any practical system based on this paper allow for such an idea. Even if the possibility is unlikely, any process short of a total recount could occasionally miss fraud obvious to a human. In addition, candidates skeptical towards the audit process have an alternative route of uncovering fraud. Appel explains and motivates this idea further.

CONCLUSION

A well-designed audit process assures the public that an election's outcome is the product of voters' intentions, not fraud or voting machine flaws. By adding a novel machine-assisted recount procedure to ballot-based audits, we can enjoy the efficiency benefits of those audits while avoiding privacy concerns and retaining the security benefits of combined paper/electronic solutions. The tests using data from Virginia's November 2006 elections confirm the efficiency advantages of machine-assisted audits, and the extended techniques that we propose promise to reduce even further the number of ballots that need to be inspected by humans.

The foregoing description of the preferred embodiment of the invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed, and modifications and variations are possible in light of the above teachings or may be acquired from practice of the invention. The embodiment was chosen and described in order to explain the principles of the invention and its practical application to enable one skilled in the art to utilize the invention in various embodiments as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto, and their equivalents. The entirety of each of the aforementioned documents is incorporated by reference herein.

What is claimed is:

1. A method for auditing ballots cast in an election, wherein each said ballot is associated with one of a plurality of voting machines and one of a plurality of groups of said ballots wherein each said group has an associated initial ballot tally, and each said group comprises all of said ballots associated with at least one of said voting machines, the method comprising the steps of:

identifying a subset of said plurality of said groups from which a plurality of sample ballots will be chosen, wherein said subset comprises fewer than all of said groups, wherein said step of identifying a subset comprises the steps of:

estimating a proportion of ballots associated with previously reported groups to a number of total anticipated ballots cast;

estimating a minimum number of sample ballots for verification;

selecting a plurality of preliminary sample ballots from said previously reported groups wherein the number of said preliminary sample ballots selected is greater than or equal to the product of said estimated proportion of ballots and said estimated minimum number of sample ballots; and

identifying all of said previously reported groups of ballots having at least one of said preliminary sample ballots;

printing an identifier on each ballot in each group of ballots in said subset of groups, wherein each ballot in each group has a different identifier than every other ballot in the same group;

performing a machine re-count of ballots in a group of ballots in said subset;

comparing results of said machine re-count to said initial tally of ballots associated with said re-counted group; and

flagging said re-counted group as containing an error if said machine re-count for said re-counted group does not equal said initial tally associated with said re-counted group.

2. A method according to claim **1**, further comprising the step of determining whether said identifiers were printed correctly on said ballots.

3. A method according to claim **1**, wherein said re-counted group comprises all ballots associated with one particular voting machine.

4. A method according to claim **1**, wherein said re-counted group comprises all ballots associated with one particular voting precinct.

5. A method according to claim **1**, further comprising the step of identifying a plurality of groups of ballots in said subset from each of which at least one ballot will be selected for manual verification.

6. A method according to claim **1**, further comprising the step of selecting a plurality of ballots for manual verification.

7. A method according to claim **6**, further comprising the step of performing manual verification of said selected plurality of ballots.

8. A method according to claim **1**, wherein said identifier comprises a serial number.

9. A method according to claim **1**, wherein said step of identifying a subset of said plurality of said groups is performed before all voting precincts have reported their votes.

10. A method according to claim **1**, further comprises the steps of:

computing a true minimum number of sample ballots after completion of reporting from all groups;

randomly selecting from all ballots a number of sample ballots equal to said true minimum number of sample ballots;

comparing how many of the randomly selected sample ballots are associated with said previously reported groups with said number of preliminary sample ballots;

if said number of randomly selected sample ballots associated with said previously reported groups is greater than said number of preliminary sample ballots, randomly selecting from said previously reported groups an additional number of sample ballots equaling a difference between said number of randomly selected sample ballots associated with said previously reported groups and said number of preliminary sample ballots; and

if said number of randomly selected sample ballots associated with said previously reported groups is less than or equal to said number of preliminary sample ballots, performing no further verifications in said previously reported groups.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO.	: 8,033,463 B2
APPLICATION NO.	: 11/833955
DATED	: October 11, 2011
INVENTOR(S)	: Edward W. Felten, Joseph A. Calandrino and J. Alex Halderman

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

ON THE TITLE PAGE:
ITEM (75) Inventors, change "Calendrino" to "Calandrino"

Signed and Sealed this
Twenty-ninth Day of November, 2011

David J. Kappos

David J. Kappos
Director of the United States Patent and Trademark Office

Verified Voting Foundation: Principles for New Voting Systems

Any new voting system should conform to the following principles:

1. It should use human-readable marks on paper as the official record of voter preferences and as the official medium to store votes.[i]
2. It should be accessible to voters with disabilities, and in all mandated languages.[ii]
3. It should provide voters the means and opportunity to verify that the human-readable marks correctly represent their intended selections, before casting the ballot.[iii]
4. It should preserve vote anonymity: it should not be possible to link any voter to his or her selections, when the system is used appropriately. It should be difficult or impossible to compromise or waive voter anonymity accidentally or deliberately.[iv] No voter should be able to prove how he or she voted.[v]
5. It should export contest results in a standard, open, machine-readable format.[vi]
6. It should be easily and transparently auditable at the ballot level. It should:
    a. export a cast vote record (CVR) for every ballot,
    b. in a standard, open, machine-readable format,
    c. in a way that the original paper ballot corresponding to any CVR can be quickly and unambiguously identified, and *vice versa*.[vii]
7. It should use commercial off-the-shelf (COTS) hardware components and open-source software (OSS) in preference to proprietary hardware and proprietary software, especially when doing so will reduce costs, facilitate maintenance and customization, facilitate replacing failed or obsolete equipment, improve security or reliability, or facilitate adopting technological improvements quickly and affordably.[viii]
8. It should be able to create CVRs from ballots designed for currently deployed systems[ix] and it should be readily configurable to create CVRs for new ballot designs.[x]
9. It should be sufficiently open[xi] to allow a competitive market for support, including configuration, maintenance, integration, and customization.
10. It should be usable by election officials: they should be able to configure, operate, and maintain the system, create ballots, tabulate votes, and audit the accuracy of the results without relying on external expertise or labor, even in small jurisdictions with limited staff.

---

[i] The medium might be voter-marked paper, a paper ballot marked with a ballot-marking device, or paper printed with the voter's selections. For the purpose of recounts or audits, the human-readable marks

[ii] This can be accomplished by providing an accessible ballot marking device.

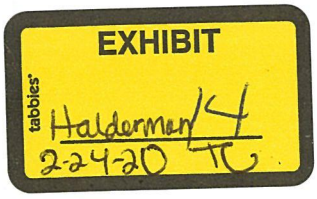
[iii] Some voters might need to rely on assistive technology, but to the extent possible, verification should not require technology.

[iv] Reporting vote subtotals by geography to comply with jurisdictional rules may entail some unavoidable loss of complete anonymity.

[v] This is to avoid coercion and vote selling.

[vi] For instance, results might be reported in EML.

1 February 2015



[vii]This might involve printing unique identifiers on ballots, if that can be done in a way that precludes linking any ballot to any to individual voter.

[viii]This includes supplies and "consumables," such as paper and batteries. Software should be licensed under a permissive license, such as BSD or MIT. Software that is not open-source should be disclosed-source to the extent reasonably possible. Disclosing source code provides the possibility of discovering errors, security vulnerabilities, and threats to voter anonymity, and mitigating their consequences. Moreover, subject to intellectual property laws, disclosing source code may offer continuity if a vendor goes out of business.

[ix]This allows modular replacement of components of a voting system. For instance, the tabulation component could be replaced without replacing the entire election management system.

[x]New ballot designs might be required for new voting methods; ballot layouts might be improved to be more usable by voters or tabulated more reliably by machine.

[xi] In particular, the design, data formats, and programming interfaces should be open and the licensing should be permissive.

Can Voters Detect Malicious Manipulation of Ballot Marking Devices?

Matthew Bernhard, Allison McDonald, Henry Meng, Jensen Hwa, Nakul Bajaj*, Kevin Chang, J. Alex Halderman

University of Michigan *The Harker School

Abstract—**Ballot marking devices (BMDs) allow voters to select candidates on a computer kiosk, which prints a paper ballot that the voter can review before inserting it into a scanner to be tabulated. Unlike paperless voting machines, BMDs provide voters an opportunity to verify an auditable physical record of their choices, and a growing number of U.S. jurisdictions are adopting them for all voters. However, the security of BMDs depends on how reliably voters notice and correct any adversarially induced errors on their printed ballots. In order to measure voters' error detection abilities, we conducted a large study (N = 241) in a realistic polling place setting using real voting machines that we modified to introduce an error into each printout. Without intervention, only 40% of participants reviewed their printed ballots at all, and only 6.6% told a poll worker something was wrong. We also find that carefully designed interventions can improve verification performance. Verbally instructing voters to review the printouts and providing a written slate of candidates for whom to vote both significantly increased review and reporting rates—although the improvements may not be large enough to provide strong security in close elections, especially when BMDs are used by all voters. Based on these findings, we make several evidence-based recommendations to help better defend BMD-based elections.**

I. Introduction

The threat of election hacking by hostile nations has prompted a major push to ensure that all voting systems in the United States have voter-verifiable paper trails, a defense recommended by the National Academies [36], the Senate Select Committee on Intelligence [53], and nearly all election security experts. Guided by past research [8], some states and localities are implementing paper trails by deploying ballot-marking devices (BMDs). In these systems, the voter makes selections on a computer kiosk, which prints a paper ballot that the voter can review before inserting it into a computer scanner to be counted [56]. BMDs have long been used as assistive devices for voters with disabilities, and a growing number of jurisdictions are purchasing them for use by all voters [24], [25], [37].

BMDs have the potential to provide better security than direct-recording electronic voting machines (DREs), which maintain the primary record of the voter's selections in a computer database and often lack a voter-verifiable paper trail. Numerous studies have demonstrated vulnerabilities in DREs that could be exploited to change election results (e.g., [11], [23], [31], [35]). In contrast, BMDs produce a physical record of every vote that can, in principle, be verified by the voter and manually audited by officials to confirm or correct the initial electronic results.

However, BMDs do not eliminate the risk of vote-stealing attacks. Malware could infect the ballot scanners and change the electronic tallies—although this could be detected by rigorously auditing the paper ballots [50]—or it could infect the BMDs themselves and alter what gets printed on the ballots. This latter variety of cheating cannot be detected by a post-election audit, since the paper trail itself would be wrong, and it cannot be ruled out by pre-election or parallel testing [51]. Instead, BMD security relies on voters themselves detecting such an attack. This type of human-in-the-loop security is necessary in many systems where detection and prevention of security hazards cannot be automated [18]. However, as several commentators have recently pointed out [7], [20], [51], its effectiveness in the context of BMDs has not been established.

Whether such a misprinting attack would succeed without detection is highly sensitive to how well voters verify their printed ballots. Every voter who notices that their ballot is misprinted and asks to correct it *both* adds to the evidence that there is a problem *and* requires the attacker to change an additional ballot in order to overcome the margin of victory. Consider a contest with a 1% margin in which each polling place has 1000 voters. If voters correct 20% of misprinted ballots, minimal outcome-changing fraud will result in an average of 1.25 voter complaints per polling place—likely too few to raise alarms. If, instead, voters correct 80% of misprinted ballots, polling places will see an average of 20 complaints, potentially prompting an investigation. (We model these effects in Section V.) Despite this sensitivity, voters' BMD verification performance has never before been experimentally measured.

In this paper, we study whether voters can play a role in BMD security. We first seek to establish, in a realistic polling place environment, the rates at which voters attempt to verify their printed ballots and successfully detect and report malicious changes. To measure these, we used real touch-screen voting machines that we modified to operate as malicious BMDs. We recruited 241 participants in Ann Arbor, Michigan, and had them vote in a realistic mock polling place using the ballot from the city's recent midterm election. On every ballot that our BMDs printed, one race was changed so the printout did not reflect the selection made by the participant.

We found that, absent interventions, only 40% of participants reviewed their printed ballots at all, only 6.6% reported the error to a poll worker, and only 7.8% correctly identified it on an exit survey. These results accord with prior studies that found poor

voter performance in other election security contexts, such as DRE review screens [1], [15] and voter-verifiable paper audit trails (VVPATs) [48]. The low rate of error detection indicates that misprinting attacks on BMDs pose a serious risk.

The risks notwithstanding, BMDs do offer practical advantages compared to hand-marked paper ballots. They allow voters of all abilities to vote in the same manner, provide a more user-friendly interface for voting, and more easily support complex elections like those conducted in multiple languages or with methods such as ranked choice [44]. BMDs also simplify election administration in places that use vote centers [56], which have been shown to reduce election costs and lower provisional voting rates [28], [42], as well as in jurisdictions that employ early voting, which can improve access to the ballot [30].

Given these advantages and the fact that BMDs are already in use, the second goal of our study was to determine whether it might be possible to boost verification performance through procedural changes. We tested a wide range of interventions, such as poll worker direction, instructional signage, and usage of a written slate of choices by each voter.

The rate of error detection varied widely with the type of intervention we applied, ranging from 6.7% to 86% in different experiments. Several interventions boosted review rates and discrepancy reporting. Verbally encouraging participants to review their printed ballot after voting boosted the detection rate to 14% on average. Using post-voting verbal instructions while encouraging participants to vote a provided list of candidates raised the rate at which voters reported problems to 73% for voters who did not deviate from the provided slate.

These findings suggest that well designed procedures can have a sizable impact on the real-world effectiveness of voter verification. We make several recommendations that election officials who already oversee voting on BMDs can employ immediately, including asking voters if they have reviewed their ballots before submission, promoting the use of slates during the voting process, informing voters that if they find an error in the printout they can correct it, and tracking the rate of reported errors. Our recommendations echo similar findings about the most effective ways to alert users to other security hazards (i.e., in context [12] and with active alerts [21]) and redirect them to take action.

Although our findings may be encouraging, we strongly caution that much additional research is necessary before it can be concluded that any combination of procedures actually achieves high verification performance in real elections. Until BMDs are shown to be effectively verifiable during real-world use, the safest course for security is to prefer hand-marked paper ballots.

Road Map Section II provides more background about human factors and security and about previous work studying the role of voter verification in election security. Section III describes our experimental setup, voting equipment, and study design. Section IV presents our results and analyzes their significance. Section V provides a quantitative model for BMD verification security. Section VI discusses the results, avenues for future work, and recommendations for improving the verifiability of BMDs. We conclude in Section VII.

II. Background and Related Work

A. Human-Dependent Security

Elections fundamentally depend on having humans in the loop—as Stark [51] notes, the voter is the *only one* who knows whether the ballot represents their intended vote—and the success or failure of election security has the potential to have history-altering effects. The type of risk posited by Stark, wherein voters do not check their paper ballots to ensure the BMD has correctly represented their selections, is a post-completion error [14], in which a user makes a mistake (or fails to verify the correctness of something) *after* they have completed the main goal of their task. Voters who forget or do not know to verify the correctness of a paper ballot after they have entered their selections on a BMD miss a critical step in ensuring the accuracy of their vote. We therefore explore how to communicate this risk to voters.

Cranor [18] describes five ways that designers can communicate risk to a user who needs to make security decisions:

1) *Warnings*: indication the user should take immediate action
2) *Notices*: information to allow the user to make a decision
3) *Status indicators*: indication of the status of the system
4) *Training*: informing users about risks and mitigations before interaction
5) *Policies*: rules with which users are expected to comply

Implementing indicators that reveal meaningful information to voters about the security status of a BMD would be next to impossible, as security issues are often unknown or unforeseen to the operators. Although voter education about the importance of verification might be an effective form of training, significant coordination would be necessary to enact such a scheme at scale. Therefore, we focus in this study on the effectiveness of warnings issued through poll worker scripts and polling place signage.

A warning serves two purposes: to alert users to a hazard, and to change their behavior to account for the hazard [62]. There are many barriers to humans correctly and completely heeding security warnings. Wogalter proposes the Communication-Human Information Processing (C-HIP) Model [61] to systematically identify the process an individual must go through for a warning to be effective. The warning must capture and maintain attention, which may be difficult for voters who are attempting to navigate the voting process as quickly as possible. Warnings must also be comprehensible, communicate the risks and consequences, be consistent with the individual's beliefs and attitudes toward the risk, and motivate the individual to change—all of which are substantial impediments in an environment with little to no user training and such a broad user base as voting.

To maximize effectiveness, warnings should be contextual, containing as little information as necessary to convey the risk and direct individuals to correct behavior [12], [61]. Voters are essentially election security novices; Bravo-Lillo et al. [12] found that, in the context of computer security, advanced and novice users respond to warnings differently. Most significantly, novice users assessed the hazard *after* taking action, whereas

advanced users assessed the hazard *before* engaging in the activity.

There may be effective ways to improve voter verification performance. Many studies have applied lessons from Cranor, Wogalter, and Bravo-Lillo et al. to help humans make secure choices in different contexts, including phishing [21], [41], browser warnings [2], [46], [52], app permissions [3], [40], and operating system interfaces [13]. In the context of phishing warnings, for example, Egelman et al. [21] found that users were far more likely to heed an active warning, or a warning that disrupted their workflow, than a passive warning. This suggests that similar interventions applied in a polling place may have a significant effect on voters' ability to review and verify their BMD ballots.

Our study contributes to this literature by exploring the effects of several modalities of warnings (oral and visual) on human detection of malicious ballot modification.

B. Voter-Verifiable Paper and Ballot-Marking Devices

A guiding principle in election security is that voting systems should be *software independent* [47]: that is, any software errors or attacks that change the reported election outcome should be detectable. Bernhard et al. [9] note that elections backed by a voter-verifiable paper record are currently the only known way to provide robust software independence. Like BMDs, voter-verifiable paper audit trails (VVPATs) and hand-marked paper ballots are widely used in an attempt to achieve software independence. However, each poses a different set of usability and accessibility challenges.

Hand-marked paper ballots record the voter's selections without the risk of having a potentially compromised computer mediating the process. However, voters often make mistakes when filling out ballots by hand that can lead to them being counted incorrectly or ruled invalid [27]. Moreover, many voters have difficulty marking a paper ballot by hand due to a disability or a language barrier. Ballots in the U.S. are among the most complex in the world, further magnifying these difficulties [38].

VVPAT technology also suffers from noted usability, privacy, and auditability problems [26]. Most implementations consist of clunky printer attachments for DREs that are difficult for voters to read, record votes in the order in which they are cast, and use a fragile paper tape. In laboratory studies, Selker et al. [48] and de Jong et al. [19] found that voters frequently did not review the VVPAT, with Selker finding that only 17% of voters detected changes between the selections they made on the DRE and those printed on the VVPAT. While there has been some criticism of Selker's findings and methodology [45], [49], their results broadly comport with work by Campbell et al. [15] and Acemyan et al. [1] about voters' ability to detect errors introduced in DRE review screens. The latter found that only 12–40% of participants successfully detected such errors.

In part due to the concerns raised by these studies, BMDs have become a popular choice for new voting system deployments in the United States. South Carolina and Georgia, together comprising nearly 9 million voters, recently adopted BMDs statewide [24], [25], as have several counties and cities, including Los Angeles County, the largest single election jurisdiction in the U.S. [58].

There has been vigorous debate among election security experts as to whether BMDs can provide software-independence (e.g., [7], [20], [51], [60]). However, the discussion has yet to be informed by rigorous experimental data. Our work seeks to fill that gap by contributing the first human-subjects study to directly measure the verification performance of voters using BMDs under realistic conditions and with a variety of potential procedural interventions.

III. Materials and Methods

Our goals in this work were to empirically assess how well voters verify BMD ballots and whether there are steps election officials can take that will enhance verification performance. To these ends, we conducted a between-subjects study where we tested several hypotheses in a simulated polling place, following the best practices recommended by Olembo et al. [39] for election human-factors research. The study design was approved by our IRB.

We sought to answer several questions, all of which concern the rate at which voters are able to detect that a BMD-printed ballot shows different selections than those the voter picked:

- What is the base rate of error detection?
- Is error detection impacted by:
  - Ballot style?
  - Manipulation strategy?
  - The manipulated race's position on the ballot?
  - Signage instructing voters to review their ballots?
  - Poll worker instructions?
  - Providing a slate of candidates for whom to vote?

In order to answer these questions in an ecologically valid way, we attempted to create an environment that closely resembled a real polling place. Nevertheless, it is impossible for any experiment to fully recreate what is at stake for voters in a real election, and so study participants may have behaved differently than voters do in live election settings. We went to extensive lengths to mitigate this limitation, and we find some data to support that we did so successfully (see Section VI-A). We used real (though modified) voting machines, printers and paper stock from deployed BMD systems, a ballot from a real election, and ballot styles from two models of BMDs. We conducted the study at two city library locations, one of which is used as a polling place during real elections.

A. The Polling Place

To provide a realistic voting experience, we structured our simulated polling place like a typical BMD-based poll site. Three investigators served as poll workers, following the script in Appendix A. Library patrons who were interested in voting began at a check-in table, where they were greeted by Poll Worker A and asked to sign an IRB-approved consent form. Participants were told they would be taking part in "a study about the usability of a new type of voting machine" and instructed

Fig. 1: *Polling Place Setup.* We established mock polling places at two public libraries in Ann Arbor, Michigan, with three BMDs (*left*) and an optical scanner and ballot box (*right*). Library visitors were invited to participate in a study about a new kind of election technology. The BMDs were DRE voting machines that we modified to function as malicious ballot marking devices.

on how to use the equipment, but they were not alerted that the study concerned security or that the BMDs might malfunction.

Each participant received a voter access card with which to activate a BMD and was free to choose any unoccupied machine. There were three identical BMDs, as shown in Figure 1. On the last day of the study, one machine's memory became corrupted, and it was removed from service; all votes that day were recorded on the other two machines.

The BMDs displayed contests in a fixed order, and voters made selections using a touch screen interface. After the last contest, the machines showed a review screen that accurately summarized the voter's selections and highlighted any undervotes. The voter could return to any contest to change the selections. A "Print Ballot" button ended the voting session and caused a printer under the machine to output the paper ballot.

Participants carried their ballot across the polling place to the ballot scanner station, where they inserted them into an optical scanner that deposited them into a ballot box. Poll Worker B was stationed by the scanner and offered instructions if necessary. Next, the poll worker collected the voter access card and asked each participant to complete an exit survey using a laptop next to the scanning station. The survey was anonymous, but responses were keyed so that we could associate them with the voter's on-screen selections, their printed ballot, and poll worker notes.

Poll Worker C, positioned separately from the other stations, acted as an observer. They verified that participants moved through the polling place stations sequentially, noted whether they spent time reviewing their printed ballots, and recorded whether they appeared to notice any abnormalities. The observer was also tasked with noting participant behavior, specifically how the participants completed each step in the voting process and any comments they made. The observer was available to answer participant questions and was frequently the poll worker participants approached upon noticing a discrepancy.

Like in a real polling place, multiple participants could progress through the voting process simultaneously. Occasionally a one- or two-person line formed as participants waited to use the BMDs or the ballot scanner.

B. The Voting Machines

BMD voting systems are currently produced by several voting machine manufacturers, the largest of which is ES&S. Over a six month period, we repeatedly attempted to engage ES&S in discussions about acquiring samples of their equipment for this study. However, these attempts were ultimately not fruitful.

Instead, we utilized AccuVote TSX DRE voting machines, which we purchased on eBay and modified to function as BMDs. The TSX was first produced by Diebold in 2003 and is still widely deployed today. At least 15 states plan to use it in at least some jurisdictions in November 2020 [57].

The TSX runs Windows CE and is designed to function as a paperless DRE or a VVPAT system. We developed software modifications that allow it to print ballots in multiple styles using an external printer. This effectively converts the TSX into a BMD—and one we could easily cause to be dishonest—while preserving the original touch-screen interface used by voters.

In order to modify the machine, we built on techniques used by Feldman et al. [23]. We began by patching the firmware so that, when the machine boots, it attempts to execute a program provided on an external memory card. We used this functionality to launch a remote access tool we created, which allowed us to connect to the TSX over a network and perform file system operations, run applications, and invoke a debugger.

The TSXes in our polling place were connected to an Ethernet switch using PCMCIA network adapters. A Python program, running on a computer on the same network, used the remote access tool's API to poll each machine for newly voted ballots. Whenever a ballot was cast, the program parsed the selections, generated a PDF file based on them, and sent it to a printer located underneath the appropriate voting machine. The program could be configured to apply different ballot styles and cheating strategies, depending on the experiment.

For every ballot, the program randomly selected one race to manipulate. In most experiments, selections could be changed in three ways: deselection in a voted-for race, selection in an unvoted-for race, or changing a selection to a different candidate. We ensured that some alteration would take place on every ballot. For example, in a vote-for-one race where the voter had made a selection, the algorithm would choose uniformly from the set of unselected choices plus no selection. One experiment used a different strategy, in which choices could only be deselected.

Both the voter's original selections and the manipulated ballot were logged for later analysis. Each voting session was associated with a unique tracking number, which was printed on the ballot along with a timestamp and encoded as a barcode.

As the final step in the voting process, participants fed their printed ballots into an AccuVote OS optical scanner, a device used to tabulate votes in parts of 20 states [57]. The scanner was intended to add realism to the experiment, but AccuVote OSes are not capable of actually tabulating the ballot styles we used. Therefore, we modified the scanner so that it simply fed each ballot into the ballot box without counting it.

We mounted a barcode reader in a 3-D printed case above the scanner's input tray and positioned it so that it would detect the ballot's tracking barcode. (This setup can be seen in Figure 3.) When the barcode was read, a Raspberry Pi would activate the AccuVote OS's feed motor to pull the ballot into the ballot box. The Raspberry Pi also displayed the ballot tracking number so that poll workers could associate the ballot with the participant's exit survey response and the observer's notes.

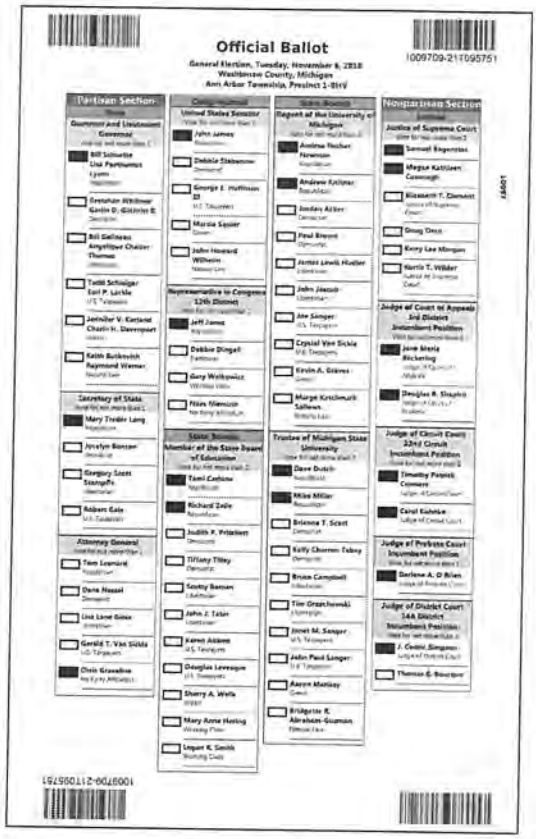
C. The Ballot

In order to ensure a realistic voting experience and increase participants' psychological investment in the outcome of the mock election, we used races and candidates from the city's actual ballot for the recent 2018 midterm election. For simplicity, we reduced the ballot to the first 13 races so that ballots would not require duplex printing or multiple pages.

We tested two ballot styles, which are illustrated in Figure 2. One is a regular ballot that shows the entire set of candidates in every race. The other is a summary ballot, which shows only the voter's selections or "NO SELECTION" if a choice is left blank. Most BMDs print ballots that resemble these styles.

The specific visual designs we used mimic ballots produced by two models of BMDs manufactured by Hart InterCivic, which also makes the voting equipment used in Ann Arbor. The regular style is also the same design as the hand-marked paper ballots most Ann Arbor voters use, ensuring that many participants found it familiar. These designs are used in jurisdictions that collectively have over 10 million registered voters [57].

The model of laser printer we used, Brother HL-2340, is certified for use with Clear Ballot's ClearAccess BMD system [43], so we chose paper stock that meets the specifications for ClearAccess [16]. Summary ballots were printed on regular weight 8.5×11 inch letter paper, while regular ballots were printed on Vellum Bristol stock 67 pound 8.5×14 inch paper.



(a) Regular Ballot

(b) Summary Ballot

Fig. 2: *Ballot Styles.* We tested two ballot styles: (*a*) a regular style, resembling a hand-marked ballot; and (*b*) a summary style, listing only the selected candidates. Both had 13 races from the city's recent midterm election. In one race, determined randomly, the printed selection differed from the voter's choice.

D. Participants and Recruitment

To gather subjects for our study, we approached staff at the Ann Arbor District Library (AADL), who offered space for us to set up our mock precinct. We conducted a total of three days of data collection in July and September 2019 at two library locations: the Downtown and Westgate branches. The Downtown branch, where our study was held for two of the three days, is an official polling location during real elections.

The AADL advertised our study through its social media feeds and offered incentives to patrons for their participation, such as points for a scavenger hunt competition [5] and souvenir flashlights [6]. We also set up a fourth voting machine outside of the mock precinct where kids could vote in an election for mayor of the library's fish tank.[1] Results from that machine were not used as part of this study, but it served as a recruitment tool for parents visiting the library with their children. In addition, we verbally recruited patrons who happened to be at the libraries during our study, using the script in Appendix B.

Participants were required to be at least 18 years of age and to sign an IRB-approved consent form. All data collected, including survey responses and behavioral observations, was completely anonymous. We informed participants that they were not required to vote their political preferences.

E. Experiments

To explore what factors affect voter verification performance, we devised nine experiments to run between subjects. In all experiments, for every participant, one selection that the participant made on the BMD was not accurately reflected on the printed ballot. Every participant within an experiment received the same instructions from the poll workers, following the script and variants in Appendix A.

The first three experiments were designed to measure verification in the absence of protective interventions. They varied the ballot style and manipulation strategy:

***E1**: Regular ballots* We used the regular ballot style and the default manipulation strategy, in which a selection could be switched, deselected, or selected if left blank by the voter.

***E2**: Summary ballots* We used the summary ballot style and the default manipulation strategy. As discussed in Section IV, we found no significant difference in error detection between regular ballots and summary ballots, so all subsequent experiments used summary ballots.

***E3**: Deselection only* To assess the sensitivity of voters to the way their ballots were changed, we limited the manipulation to deselecting one of the voter's choices at random.

Four further experiments tested interventions to determine if they improved error detection. We tried posting a sign and having poll workers give different instructions at various times:

***E4**: Signage* A sign was placed above the scanner that instructed voters to check their printed ballots, as shown in Figure 3. We designed the sign following guidelines from the U.S. Election Assistance Commission [55].

Fig. 3: *Warning Signage.* One of the interventions we tested was placing a sign above the scanner that instructed voters to verify their ballots. Signage was not an effective intervention.

***E5**: Script variant 1* During voter check in, the poll worker added this instruction: "Please remember to check your ballot carefully before depositing it into the scanner."

***E6**: Script variant 2* When the voter approached the scanner, the poll worker said: "Please keep in mind that the paper ballot is the official record of your vote."

***E7**: Script variant 3* When the voter approached the scanner, the poll worker said: "Have you carefully reviewed each selection on your printed ballot?"

The final two experiments assessed whether reminding participants of their selections during verification improved their performance. We gave voters a slate of candidates for whom to vote that they could carry with them throughout the voting experience. While we refer to this as a slate, a sample ballot that the voter filled in before voting could serve the same purpose. Every voter received the same slate (Appendix C), which was randomly generated and contained an even mix of parties.

***E8**: Slate with script variant 2* Voters were given the slate. Poll workers encouraged verification with script variant 2.

***E9**: Slate with script variant 3* Voters were given the slate. Poll workers encouraged verification with script variant 3.

[1] Mighty Trisha unexpectedly beat Creepy Bob, leading some Bob supporters to complain that the results were fishy [4].

Experiment	N	Were observed examining ballot	Reported error on exit survey	Reported error to poll worker
Without interventions:				
E1: Regular ballots	31	41.9%	6.5%	6.5%
E2: Summary ballots	31	32.3%	6.5%	6.5%
E3: Deselection only	29	44.8%	10.3%	6.9%
Subtotal/Mean	91	39.7%	7.8%	6.6%
With interventions:				
E4: Signage	30	13.3%	3.3%	6.7%
E5: Script variant 1	30	46.7%	13.3%	6.7%
E6: Script variant 2	25	92.0%	16.0%	16.0%
E7: Script variant 3	31	38.7%	19.4%	12.9%
E8: Slate with script variant 2	13	100.0%	38.5%	38.5%
E9: Slate with script variant 3	21	95.2%	71.4%	85.7%
Subtotal/Mean	150	64.3%	24.0%	27.8%

TABLE I: *Verification Performance for Each Experiment.* Without interventions, participants' verification performance was remarkably poor: only 7.8% noted on an exit survey that their ballots had been altered, and only 6.6% informed a poll worker (averaged across experiments). The various interventions we tested had widely different effects, ranging from no significant improvement (**E4**, **E5**) to a large increase in verification success (**E8**, **E9**).

IV. Results

A. Participant Demographics

We recruited 241 participants. The vast majority (220, 91%) indicated that they were native English speakers; 19 reported speaking twelve other native languages, including Hungarian, Korean, and Arabic; and two subjects gave no response. Participants who disclosed their age ranged from 18 to 84 years old, with a mean of 43.7 and a median of 42; 15 subjects did not answer the question. The percentages that follow are out of the total number of responses to each question: Respondents identified as male (84, 35%), female (152, 64%), or other (3, 1%); two did not respond. Subjects reported their ethnicity as Caucasian (187, 80%), Asian (17, 7%), African American (6, 3%), Mexican American/Chicano (5, 2%), and Other Hispanic/Latino (9, 4%); others reported not having any of these ethnic backgrounds (2, 1%) or were multiracial (9, 4%). Participants reported their level of educational attainment as some high school (1, 0.4%), a high school diploma (4, 2%), some college (20, 8%), a two-year degree (10, 4%), a four-year degree (80, 33%), a master's or professional degree (92, 38%), or a doctorate (34, 14%).

Most subjects indicated that they were registered to vote in the U.S. (220, 92%), had voted in a previous election (216, 91%), and had voted in the November 2018 midterm election (209, 87%). However, we note that, historically, 38–45% of non-voters have been found to falsely report having voted [10].

Compared to the population of Ann Arbor at the time of the 2010 census, our participant pool overrepresented Caucasians ($\Delta = 7.6\%$) and underrepresented African Americans ($\Delta = -4.4\%$) and Asians ($\Delta = -8.7\%$) [54]. The study population also overrepresented females ($\Delta = 13\%$) and underrepresented males ($\Delta = -16\%$) [59]. In other reported aspects, participants' demographics resembled the population of Ann Arbor voters (the city is among the most highly educated in the U.S.) [33].

B. Verification Performance

To quantify verification performance, we collected three data points for each participant, which are summarized in Table I. First, an observer noted whether the subject appeared to examine the printed ballot for at least two seconds. Second, the exit survey asked, "Did you notice anything odd about your ballot?", and we recorded whether the subject's response corroborated the discrepancy (i.e., correctly articulated which race was changed). Third, we recorded whether subjects reported the ballot modification to a poll worker. Most experiments saw more participants identify discrepancies in the survey than were reported to poll workers, but these differences were not statistically significant. Where applicable, we refer to participants who by some means reported detecting the discrepancies as "noticers" and those who did not as "non-noticers".

*1) Performance without interventions (**E1–E3**):* With no interventions, we found verification performance to be consistently poor. The three experiments involved 91 participants, and, averaged across the experiments, only 40% of participants examined their ballots, only 7.8% noted the error on the exit survey, and only 6.6% reported it to a poll worker. We did not find significant differences in performance between regular and summary ballots or between the tested attack strategies.

*2) Effectiveness of interventions (**E4–E9**):* The tested interventions resulted in a wide range of effect sizes. Neither signage (**E4**) nor poll worker instructions issued before the participant began voting (**E5**) yielded a statistically significant improvement to any aspect of verification performance. In

contrast, poll worker instructions issued *after* the ballot was printed (**E6** and **E7**) did have a positive effect, boosting reporting rates to 20% on the exit survey and 14% to poll workers (averaged across the experiments).

The largest performance gains occurred when participants were directed to vote using a slate of candidates (**E8** and **E9**). However, only **E9** produced a statistically significant difference in reporting rates (Fisher's exact $p < 0.001$).[2] Averaged across both experiments, reporting rates increased to 55% on the exit survey and 62% to poll workers. **E8**, in which participants were directed how to vote using a slate of candidates, saw detection and reporting rates of 39%, which is similar to results for DRE review screen performance found by Campbell et al. [15] and Acemyan et al. [1], in studies that similarly directed participants how to vote. With script variant 3, the use of a slate produced a significant difference (comparing **E7** and **E9**, Fisher's exact $p < 0.02$) for both review and report, but it did not produce a significant difference using script variant 2 (comparing **E6** and **E8**). This indicates that voters may be sensitive to the specific instructions they receive about reviewing their ballots.

C. Correlates

1) Reviewing the ballot: Reviewing the ballot at all was significantly correlated with error reporting (two-sample permutation test $p < 0.001$ with 10k repetitions). Some interventions do seem to promote reviewing: **E6**, **E8**, and **E9** saw significant increases (Fisher's exact $p < 0.004$), although **E7** did not.

2) Time to ballot submission: Careful verification takes time, so one might expect that participants who noticed discrepancies took more time to cast their ballots. As an upper bound on how long subjects spent verifying, we calculated the time from ballot printing to ballot submission. (Due to clock drift on one of our machines, data from the third day of experiments was unusable, and consequently **E4** and **E7** are excluded from our timing analysis.) As expected, we find that noticers took an average of 121 s between printing and ballot submission (median 114 s), compared to only 43 s for non-noticers (median 32 s). This difference is statistically significant (two-sample permutation test $p < 0.004$, 10k iterations).

We compared the submission times for two sets of experiments: ones with extra instructions to the voter (**E5**, **E6**, **E8**, and **E9**; $N = 84$) and ones without (**E1**, **E2**, and **E3**; $N = 91$). The experiments that asked participants to review their ballots saw significantly more time spent between ballot printing and submission (two-sample permutation test $p < 0.004$, 10k iterations), an average of 83 s (median 72 s) compared to 50 s without (median 33 s).

Notably, participants who were given a slate of candidates to vote for had much higher submission times (two-sample permutation test $p < 0.004$, 10k iterations). Noticers in the slate experiments took an average of 119 s (median 111 s) and non-noticers averaged 55 s (median 52 s). This might be partly attributed to voters having to select unfamiliar candidates and wanting to check their work.

[2]All p-values were computed with a Bonferroni correction at a family-wise error rate of 0.05.

3) Demographics: Comparisons of detection rates across demographic groups revealed that a strong indicator for verification performance was voting experience. Subjects who reported being registered to vote ($N = 220$) detected errors with their ballots 19% of the time, while their those who did not ($N = 21$) detected errors 4.8% of the time. Those who reported voting previously ($N = 216$) caught ballot discrepancies in 19% of cases, again performing better than those who reported not voting before ($N = 25$), who detected an error in 4.0% of cases. If someone reported voting in the 2018 midterm election ($N = 209$), they detected problems with their ballot 20% of the time, whereas if they did not ($N = 32$), they detected problems 3.1% of the time. This may indicate that familiarity with the midterm ballot we used caused participants to feel more invested in the accuracy of their votes; however, we did not establish this to statistical significance.

Other demographic factors, such as age, education, ethnicity, and gender, had no correlation with detecting manipulation.

4) Ballot position: Noticing was correlated with ballot position (Pearson's of -0.64), indicating that discrepancies in more prominent races are more likely to be noticed. (Race 0 was the first race on the ballot, so the number of noticers decreases as the race position increases, hence the negative correlation coefficient.) On our ballot, the first five races (Governor, Secretary of State, Attorney General, U.S. Senator, and Representative in Congress) were prominent partisan contests with a high likelihood of name recognition. In the experiments with no intervention (**E1–E3**), 37 participants had one of these races manipulated, and five reported the error on the exit survey, a rate of 14%. Additional experiments are necessary to establish the strength of this effect when combined with interventions.

5) Undervotes: A metric that may inform voters' ability and willingness to verify their ballot is how much care they take in filling out the ballot. There are two metrics we use to examine this: whether a participant voted in every contest on the ballot, and whether the participant voted in every available position on the ballot (e.g., in a vote-for-two contests, the participant selected two choices). Table II shows the rates of voting in every race and every position on the ballot, with **E8** and **E9** removed as they directed participants to vote in every position. Voters who noticed discrepancies voted in every race or every position at a higher rate than those who did not, but not significantly so (likely due to our small sample size). Since these undervotes are visible to malware running on a BMD, this correlation could be exploited by an attacker to focus cheating on voters who are less likely to carefully verify, provided future work more firmly establishes this link.

	Overall	Noticers	Non-noticers
Every race	64.3%	73.9%	63.0%
Every position	43.0%	47.8%	42.4%

TABLE II: *Participant Attentiveness.* Voters who noticed the discrepancy tended to vote in every race and ballot position more often than those who did not.

6) Partisanship: To assess the role partisanship plays in detection rates, we scored each ballot with a partisanship score, where a vote for a Democratic candidate was scored -1 and a vote for a Republican candidate was scored 1, and we take the absolute value of the sum. There were 11 opportunities to vote in a partisan way, so a participant who voted straight-party for either major party would achieve a score of 11. Excluding **E8** and **E9**, where voters were directed how to vote, the mean partisanship score for our participants was 8.3, and the median was 11. Although our BMD did not offer an automatic "straight-party" voting option, 105 participants achieved the maximum partisanship score.

Intuitively, a voter expecting every selected candidate to be from the same party might be more likely to notice a selection from a different party. Looking at only these straight-party voters, 15 out of 105 detected the errors. Of those, nine had a partisan race swapped to a different candidate of a different party, and six of those participants wrote in the survey that they had detected the change based on party. For example, one participant wrote, *"voted GOP for governor / lieutenant governor but Libertarian was actually selected on the paper ballot."*

This suggests that choosing a uniform set of candidates may help voters detect when something has gone wrong on their ballot, although more work is needed to establish that this is indeed the case, especially in more politically diverse populations. If this positive effect holds, it could be further promoted with ballot designs that prominently display the party, which could help voters see the information that is important to them while they review the ballot. On the other hand, BMD malware could be designed to counter this effect by focusing cheating on voters who do not cast a straight-party ballot.

7) Slate voting: 34 participants were assigned an intervention which asked them to vote for a preselected slate of candidates (with a partisanship score of 0). Of these, only 26 participants voted exactly as directed. Of the eight participants who did not, four voted a straight Democratic ticket (partisanship score of 11), one voted a heavily Democratic ticket (score of 9), two voted slightly Democratic tickets (scores of 3 and 5), and one voted a non-partisan ticket (score of 0), which only deviated from the slate in five positions. Of the eight participants who deviated from the slate, no participant deviated by fewer than five positions, indicating that either the deviation was deliberate or our instructions to vote the slate were unclear. Only one deviating participant managed to notice the discrepancy on their ballot, leaving participants who deviated from the slate a 13% notice rate compared to the 73% notice rate for those who did not deviate.

8) Network effects: One potential feature of a live polling place environment is a network effect: will a voter who is voting at the same time as a noticer be more likely to notice a problem on theirs? However, the number of people who notice in a given experiment is a confounding factor: voters are more likely to overlap with a noticer if there are more noticers. To interrogate this, we ran partial hypothesis tests for each intervention using Fisher's exact tests with permutations of overlapping with a noticer and noticing, and then combined using Fisher's combining function. We found that the effect of overlapping with a noticer did not significantly impact whether a participant noticed. This suggests that our interventions were more important than overlapping.

9) Signage: One feature that did not correlate with improved verification performance was the signage we tested (**E4**). Our observer noted that 11 of 30 participants in the signage experiment did not notice the sign at all. Only two participants in this experiment detected the modification of their ballot and reported it, and only one accurately noted the discrepancy in their survey, suggesting that passive signage alone may be insufficient to capture voters' attention and shape their subsequent behavior.

D. Participant Comments

Participants had two free-response sections in the exit survey. The first asked about anything "odd" they had noticed about the ballot. The second invited any additional comments. Of the 241 participants, 114 responded to at least one of these prompts. We note several features of their responses.

1) Discrepancy reports: In total, 44 participants (18%) noted in the free response section of the survey that they had identified some discrepancy on their paper ballot. Of these, 31 correctly identified the change, 12 gave no detail (e.g., *"At least one of my choices did not match who I picked"*), and one incorrectly identified the change (but did report that there was a mistake). We omitted this last participant from our "noticers" category where applicable.

Of the 44 participants who reported a change on their ballot in the survey, five added that they thought it could have resulted from a mistake they made. For example, one participant reported: *"I don't remember voting for the member of Congress and there was a vote. I very well may have but just don't remember."*

2) Attitudes about verification: Twelve participants mentioned either that they would only be comfortable voting on a paper ballot or that they were comforted by the fact that a paper trail was created. Only three of these 12 participants noticed that their ballot had been modified, despite the fact that they recognized that the paper ballot was an important tool for ensuring election integrity.

Several participants seemed to realize *after* casting their vote that the evaluation of their paper ballot was important; 13 participants mentioned in the survey that they did not review or that they should have reviewed the ballot, although we did not ask them about it. This concern may have been triggered by our survey question about what they had noticed about the paper ballot, but it also might be an indication that our interventions did cause voters to think about the risk—albeit too late.

The free responses also indicate that some participants assumed that the vote was completed and submitted on the BMD, rather than the paper ballot being the official record of their vote. One participant wrote, *"I was surprised to still have a paper ballot, after using the touch system. I was expecting the results to be registered electronically."* This assumption may discourage voters from verifying the selections on their

paper ballot. Similarly, another participant, prompted by script variant 3 ("Have you carefully reviewed each selection on your printed ballot?"), responded to a poll worker, *"I checked it on the screen, it better be right."*

Three participants expressed concern that they would not know what to do if they noticed a problem with their paper ballot during a real election. One person wrote, *"Having the printout be incorrect was confusing and it's not clear how that would be handled in an election environment."*

3) Feedback on the BMDs: We told participants that the experiment was a study about a new kind of voting system, and many left feedback about the interface and appearance of the machines. In Michigan, where we conducted the study, BMDs are available in every precinct, but voters must request to use them. The vast majority of voters use hand-marked paper ballots, so study participants were likely unfamiliar with BMD voting. In their comments, 21 participants expressed liking the system, while only three disliked it. Although merely anecdotal, this reflects previous findings that voters like touch-screen voting equipment [22].

V. Security Model

We are primarily motivated by the threat of undetected changes to election outcomes due to BMD misprinting attacks. Prior work has shown that such attacks cannot be reliably ruled out by pre-election or parallel testing [51], and we seek to answer whether voter verification can be an effective defense.

If a voter reports that their printed ballot does not reflect their on-screen selections, what should election officials do? Unfortunately, there is not yet a practical way to prove that the BMD misbehaved during voting. From officials' perspective, it is also possible that the voter is mistaken, or even lying, and in a large voter population, there will always be some rate of spurious problem reports, even when BMDs are working correctly.

For these reasons, problem reports from voters can serve only as evidence that something *might* be wrong with the BMDs. If the evidence exceeds some threshold, officials could invoke contingency plans. For instance, they could remove BMDs from service to minimize further damage, perform forensic investigations in an attempt to uncover the cause, or even rerun the election if outcome-changing fraud cannot be ruled out.

Any of these responses would be costly (and none is foolproof), so the threshold for triggering them should not be too low. Moreover, attackers could exploit a low threshold by recruiting voters to fraudulently report problems, in order to disrupt or discredit the election. On the other hand, if the threshold is too high, outcome-changing fraud could be ignored.

To better understand how verification performance affects security in this setting, we construct a simple model. We assume, optimistically, that the attacker has no way to guess whether a particular voter is more likely than average to detect the alteration, and so chooses voters to attack at random. We further assume that whenever voters detect problems, they are able to remedy them and cast a correct vote by hand-marking a ballot. Except where noted, the model assumes that all voters cast their votes using BMDs.

Number of problem reports Let d be the fraction of misprinted ballots that voters detect, report, and correct. Suppose a contest had n ballots cast, and the reported fractional margin of victory was m. To have changed the outcome, the attacker would have had to successfully modify at least $n\frac{m}{2}$ cast ballots. However, since some modifications would have been corrected, the attacker would have had to induce errors in a greater number of printouts: $n\frac{m}{2(1-d)}$. Under our optimistic assumptions, if the attack changed the outcome, we would expect the fraction of voters who reported problems, a, to exceed:

$$a > m\frac{d}{2(1-d)}.$$

The model shows that the security impact of verification is non-linear, because every voter who corrects an error *both* increases the evidence that there is a problem *and* forces the attacker to cheat more in order to overcome the margin of victory. Figure 4 illustrates this effect.

With the 6.6% error detection rate from our non-intervention experiments and a close election with a 0.5% margin (the margin that causes an automatic recount in many states) a successful attack would cause as few as 0.018% of voters—less than 1 in 5000—to report a problem. Small changes in verification performance around our base rate cause relatively little change in the amount of evidence. More than doubling the error detection rate to 14% (the rate we found for prominent races) only increases the fraction of voters who report a problem to 0.039%. However, larger improvements have an outsized effect: with the 86% error detection rate from our most successful experiment, at least 1.5% of voters (1 in 67) would report problems.

Required detection rate Suppose election officials activate a countermeasure if the fraction of voters who report problems

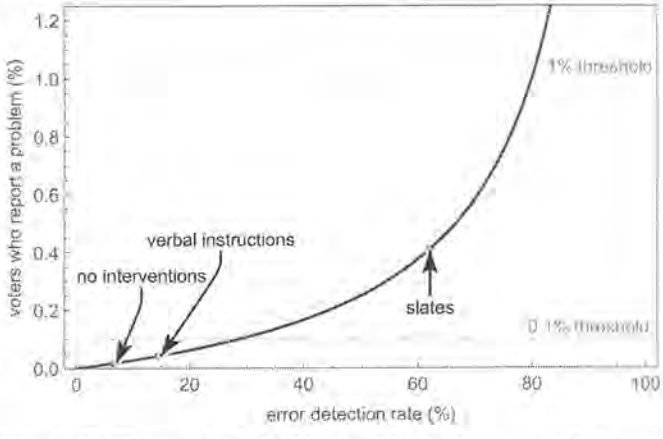


Fig. 4: *BMD security is highly sensitive to human performance.* Given a 0.5% margin of victory, we plot the percentage of voters who report a problem during the minimal outcome-changing attack as a function of the rate at which errors are detected and corrected. This model implies that using BMDs safely for all voters requires dramatically improved verification performance or very sensitive attack detection thresholds.

exceeds a threshold a^*. For a given margin, the countermeasure will be triggered by minimal outcome-changing fraud when:

$$d > \frac{2a^*}{m+2a^*}.$$

An expensive countermeasure, like rerunning an election, will require a high trigger threshold—say, 1% of voters reporting a problem—to avoid false positives. With a 0.5% margin, reaching a 1% trigger threshold would require an error detection rate exceeding 80%. A less expensive countermeasure, such as an investigation, might be triggered by a lower threshold—say, 0.1%. Reaching this lower threshold in an election with a 0.5% margin would require an error detection rate greater than 29%. This suggests that using BMDs securely for all voters will require large improvements to verification performance or extremely low thresholds for triggering countermeasures.

Minimizing BMD voting helps dramatically Securing against misprinting attacks is far easier if only a small fraction of voters use BMDs than if all in-person voters do. This is because an attacker would be forced to cheat on a much larger fraction of BMD ballots in order to achieve the same change to the election results. Moreover, if the population of BMD voters is smaller than half the margin of victory, it is impossible for a BMD misprinting attack to change the outcome.

Let b be the fraction of voters who use BMDs. We can replace m in the expression above with $\frac{m}{b}$ and let a^* be the fraction of *BMD voters* that must report a problem to trigger the countermeasure. In Maryland, which uses hand-marked paper ballots but makes BMDs available to voters who request them, 1.8% of voters use BMDs [34]. With a 0.5% margin, as in the previous example, Maryland would reach a complaint threshold of 1% of BMD voters with an error detection rate of only 6.7%. If 5% of voters use BMDs, the error detection rate would need to be 17%. Our results suggest that these more modest rates of verification likely are achievable, in contrast to the far greater accuracy required when all voters use BMDs.

This model overestimates security An attacker might use any number of features (including several of the correlations we observed) to focus cheating on voters who are less likely to successfully catch errors. For instance, an attacker could preferentially modify ballots that have undervotes or a mix of selections from different parties. Attackers could also selectively target voters with visual impairments, such as those who use large text or an audio ballot. Other features, such as how long voters spend inspecting the candidate review screen, might also prove to be predictive of verification success. For these reasons, our simplified model is likely to overestimate the effectiveness of verification against sophisticated attackers.

We also note that some attackers may merely seek to cast doubt on election results by causing highly visible errors or failures—which are also possible with hand-marked paper ballots. However, in general, BMDs are vulnerable to all classes of computer-based attacks that affect hand-marked paper ballots and to others, such as the misprinting attack discussed here, to which hand-marked paper ballots are not susceptible.

VI. Discussion

A. Limitations

It is challenging to capture real-world voter behavior in a mock election. However, our study followed established best practices [39], and we strived to create as realistic a polling environment as we could. It is impossible to know exactly how well we succeeded, but the effect seems to have been convincing: several people approached us to ask whether there was a real election taking place that they had not heard about. Our participants also seemed engaged in the study; many expressed strongly held political preferences in our survey (so much so that some refused to vote according to our slate), and a large majority reported voting in the 2018 midterm. On the other hand, the election used a ballot that was more than nine months old, which may have reduced participant motivation, and we had a few participants who reported that they did not vote in our state or were otherwise unfamiliar with our ballot. It is also possible that our results were skewed due to selection bias and observer effect.

Another limitation of our work is that we drew participants from a population that is locally but not nationally representative. Our participants tended to be younger, significantly better educated, more liberal, more likely to be female, and more likely to be Caucasian than the average voter in the United States [54]. Future work is needed to validate our study in more diverse and representative populations.

Although our results suggest that certain interventions can boost verification performance, the data is too sparse to provide a high-fidelity understanding of the magnitude of the improvements. In addition, due to time constraints, we were unable to test the interplay of all combinations of interventions, and some interventions appear to be sensitive to small changes (e.g., the difference in phrasing between script variants 2 and 3). Further study is needed to better characterize what makes interventions work and how they interact before we can confidently conclude that any particular set of procedures will be effective in practice.

B. Discussion of Findings

Our study provides the first concrete measurements of voter error detection performance using BMDs in a realistic voting environment. At a high level, we found that success rates without intervention are very low, around 6.6%. Some interventions that we tested did not significantly impact detection rates among participants, although others improved detection drastically and may serve as a roadmap for interventions to explore in further research. We discuss those interventions here.

1) Verbal instructions can improve verification: Notably, all interventions that involved poll workers verbally encouraging verification between the BMD and the scanner—those in **E6–E9**—resulted in higher ballot reviewing and error reporting rates. This, coupled with the fact that reviewing the printout was highly correlated with error detection across all of our results, suggests that interventions focused on causing the voter to review the ballot carefully may be helpful. On the

other hand, instructions at the beginning of the voting process (**E5**) and passive signage (**E4**) had no significant effect on error reporting. This pattern of effects is supported by findings from the usable security literature, which suggest that post-completion errors can be mitigated with timely interruptions that encourage individuals to take defensive steps [14].

It is worth noting that we also found that these interventions caused participants to take longer to submit their ballots, on average about twice as long. This could cause longer lines at polling places if these interventions are implemented without complementary procedural considerations, such as having adequate space for voters to stop and review their ballots.

2) Effectiveness of slates: Directing participants to vote for a provided slate of candidates, combined with verbally prompting them to review their printouts, resulted in strongly increased rate of error detection: 74% of participants who were given a slate and did not deviate from it noticed the errors. This finding may suggest that encouraging voters to write down their preferences in advance can boost verification.

However, the slates we used functioned quite differently from slates likely to be used in practice. The choices we provided were randomly generated and had no basis in the subject's preferences—in a real election, slates would reflect who the voter intended to vote for, most likely created by the voter or their political party [29]. It is possible that the success rate we observed was primarily due to participants carefully attempting to follow our instructions and vote for unfamiliar candidates. Further study is needed with more realistic slate conditions (i.e., asking subjects to write down their preferences) in order to assess whether slates really do help voters catch errors.

C. Recommendations

Since BMDs are widely used today, we recommend several strategies for improving voter verification performance. While we are unable to conclude that these strategies will enhance error detection to the point that BMDs can be used safely in close or small elections, our findings indicate that they can help.

1) Design polling places for verification: Polling place lay-out and procedures should be designed with verification in mind. As we have discussed, voters need time and space to verify their ballots. If tables or areas to stand out of the way are provided, voters will be able to carefully verify without causing lines to form or slowing polling place throughput. The presence of such a "verification station" might also encourage verification.

Another practical concern is privacy. Several of our participants expressed discomfort with the fact that we did not provide a privacy sleeve for their ballots (a requirement in Michigan), and that the scanner accepted the ballots face-up only, with one participant stating, *"I feel like inserting the ballot face up in the scanning machine will make people uncomfortable."* Voters may not feel comfortable reviewing their ballots in front of poll workers but may be unsure where to go to review them privately.

2) Incorporate post-voting verbal instructions: As all of our script-based interventions that took place after the ballot was printed (**E6–E9**) showed an increase in verification performance, we recommend that poll workers interrupt voters after their ballot has printed but before it is scanned and ask them to review it. Signage with a similar message to our scripts placed at the optical scanner (**E4**) or instructions before the participants voted (**E5**) did not result in significant differences in error detection; nevertheless, further study with additional variations is prudent before ruling out such strategies.

3) Encourage personalized slate voting: Although our study tested randomized slates, rather than personalized slates, the effect size was so large that we tentatively recommend encouraging the use of personalized slates by voters. In our experiments (**E8** and **E9**), participants who were directed to vote using a randomized slate (and did not deviate) reported errors at a rate of 73%. If voters prepare their own slates at home (or use a printed slate prepared, for instance, by a political party or other organization), they can use them to check each selection on the BMD printout. We note that, since we did not directly test the use of personalized slates, further research is necessary to ascertain whether large performance gains are actually achieved. Furthermore, even if personalized slates are effective, the gain will be limited to the fraction of voters who can be induced to use them.

Slates have potential downsides and should be used with care. They have the potential to compromise ballot secrecy, so we recommend providing a closed trash can, paper shredder, or other means for voters to privately dispose of them before leaving the precinct. Coercion is also a threat, but voters could be advised to prepare multiple different slates as a defense.

4) Help voters correct errors, and carefully track problems: Verification-promoting interventions will be of little use if action cannot be taken to remedy misbehaving BMDs—something that even our participants expressed concern about.

First, it is crucial that polling places have a procedure for voters who want to correct their printed ballots. Several subjects commented that they would not know what to do if something was wrong with their ballot in a real election, indicating that this problem is present in current election procedures.

Second, detailed records should be kept about which BMD the voter used and what the specific issue was, including the contest and candidates involved (to the extent that the voter is willing to waive ballot secrecy). Problems should be treated as potentially serious even when the voter believes they are at fault—we note that several participants in our study believed they had made a mistake even though the BMD actually was programmed to be malicious. Problem reports should be centrally reported and tracked during the election, so that issues affecting multiple precincts can be identified as rapidly as possible.

5) Prepare contingency plans: What to do in the event that BMDs are known or suspected to be misbehaving is a more difficult question. If an elevated number of voters have a problem with a single machine, it should be taken out of service, provided there are other BMDs available for use (especially for voters with disabilities, who may have no alternative).

If widespread problem reports occur—particularly problems focused on a tightly contested race or significantly exceeding the rate reported in past elections—officials could consider

taking most BMDs out of service and encouraging all remaining voters who can to use hand-marked ballots. This raises logistical challenges: polling place would need to have enough ballots available for hand-marking, or the ability to print ballots on demand, and votes already cast on the BMDs would be suspect.

After the election, forensic analysis of the BMDs could be performed to attempt to determine the cause of reported errors. Unfortunately, such analysis cannot in general rule out that a sophisticated attack occurred and left no digital traces. Even if programming errors or attacks are uncovered, they may be impossible to correct if officials are unable to determine whether the effects were large enough to change the election outcome. The only recourse might be to re-run the election.

Our findings show that, in the event of an actual error or attack, the rate of reported problems is likely to be only the tip of the iceberg. In our non-intervention experiments, undetected errors outnumbered reported problems by almost twenty to one. Our results further suggest that an attacker who cleverly focused cheating on voters who were less likely to verify could achieve an even higher ratio of undetected errors. An effective response requires either being very sensitive to reported problems—which increases the chances that an attacker could trigger false alarms—or achieving very high error correction rates.

6) Educate voters about BMD operations and risks: Like in other human-in-the-loop security contexts, greater education could boost voters' awareness of the importance of careful verification and boost error detection and reporting rates.

To this end, we recommend educating voters that the paper, rather than what the BMD screen shows, is the official record of their votes. Several of our participants said they realized after scanning that they should have, but did not, review their printouts. Others stated that they had checked the review screen on the machine and that they trusted the paper to be correct. It is likely that many participants incorrectly assumed that the BMDs, rather than the paper and scanner, tabulated their votes.

We also recommend educating voters about the possibility of BMD malfunction. Many of our participants seem not to have even considered that the machine might have changed their votes, as indicated by the voters who blamed themselves for the misprinted ballots. Raising threat awareness could help motivate voters to carefully inspect the paper, as well as give them greater confidence to report any discrepancies they detect.

7) Consider the needs of voters with disabilities: Further research is needed to specifically examine verification performance among voters with disabilities, but we offer some initial recommendations here. Detecting errors in printed ballots may be especially challenging for voters with impaired vision. Designing BMD ballots for maximum legibility might help, and so might encouraging voters who use text-to-speech devices to bring them to the polls for use during verification. Jurisdictions could also provide air-gapped accessible devices to read the ballot back to voters, in case voters do not have their own text-to-speech devices. These steps would have the added benefit of reinforcing the message that the content of the paper ballots is what gets counted. If BMDs are to live up to the promise of better and more accessible voting, enabling all voters to verify their printed ballots is a must.

8) Require risk-limiting audits: Even perfectly verified paper ballots are of little use for security if they are not rigorously audited to confirm the results of computer-based tabulation. Fortunately, risk-limiting audits [32] (RLAs) are gaining momentum in the United States. Colorado, Nevada, and Rhode Island mandate statewide RLAs, and states including Michigan, Virginia, Georgia, and Pennsylvania are considering implementing them soon [17]. RLAs and effective verification are both necessary in order for paper to provide a strong defense against vote-stealing attacks, and we recommend that efforts to achieve both be pursued vigorously.

VII. Conclusion

We conducted the first empirical study of how well voters using BMDs detect errors on their printed ballots, which is a limiting factor to the level of security that a BMD-based paper trail can provide. Based on the performance of 241 human subjects in a realistic polling place environment, we find that, absent specific interventions, error detection and reporting rates are dangerously low. Unless verification performance can be improved dramatically, BMD paper trails, particularly when used by all in-person voters, cannot be relied on to reflect voter intent if the machines are controlled by an attacker.

Nevertheless, we also find that procedural interventions can improve rates of error detection and reporting, potentially increasing the security offered by BMDs. The interventions we tested should serve as examples of what is and is not likely to be effective, and we hope they will point the way for further research and experimentation. These findings add to the broad literature of human-in-the-loop security results and recommendations, and they provide additional examples of what does and does not work in human-centric security.

Our results should not be read as demonstrating that BMDs can be used securely. Further work is needed to explore the potential for attackers to predict which voters will verify, and additional human-subjects testing is necessary to confirm whether sufficient rates of verification success can be achieved in practice. The cost of implementing interventions and contingency plans may also be prohibitive. Nevertheless, BMDs do offer advantages, including uniform accessibility and ease of administration. We hope our work will help election officials make better informed choices as they weigh these benefits against the security risks of using BMDs for all voters.

Acknowledgments

The authors are grateful to Jackie Fleischer Best, Eli Neiburger, Emily Howard, Matt Dubay, and everyone at the Ann Arbor District Library, without whom this study would not have been possible. We also thank Philip Stark for advice about our statistical analyses; Ben Adida, Monica Childers, and Ben VanderSloot for feedback about the experimental design; and the anonymous reviewers. This material is based in part upon work supported by the National Science Foundation under Grant No. CNS-1518888, by the Facebook Fellowship Program, and by the Andrew Carnegie Fellows Program.

References

[1] C. Z. Acemyan, P. Kortum, and D. Payne. Do voters really fail to detect changes to their ballots? An investigation of ballot type on voter error detection. *Proceedings of the Human Factors and Ergonomics Society*, 57:1405–1409, 2013.

[2] D. Akhawe and A. P. Felt. Alice in Warningland: A large-scale field study of browser security warning effectiveness. In *22nd USENIX Security Symposium*, pages 257–272, 2013.

[3] H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal. Your location has been shared 5,398 times! A field study on mobile app privacy nudging. In *33rd ACM Conference on Human Factors in Computing Systems*, CHI, pages 787–796, 2015.

[4] Ann Arbor District Library. Classic shop drop! Plus, fish election results!, Aug. 2019. https://aadl.org/node/396262.

[5] Ann Arbor District Library. Mock the vote, July 2019. https://aadl.org/node/395686.

[6] Ann Arbor District Library. Mock voting @ AADL, Sept. 2019. https://aadl.org/node/397364.

[7] A. Appel, R. DeMillo, and P. Stark. Ballot-marking devices (BMDs) cannot assure the will of the voters, 2019. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3375755.

[8] S. Bell, J. Benaloh, M. D. Byrne, D. DeBeauvoir, B. Eakin, G. Fisher, P. Kortum, N. McBurnett, J. Montoya, M. Parker, O. Pereira, P. B. Stark, D. S. Wallach, and M. Winn. STAR-Vote: A secure, transparent, auditable, and reliable voting system. *USENIX Journal of Election Technology and Systems*, 1(1), 2013.

[9] M. Bernhard, J. Benaloh, J. A. Halderman, R. L. Rivest, P. Y. Ryan, P. B. Stark, V. Teague, P. L. Vora, and D. S. Wallach. Public evidence from secret ballots. In *2nd International Joint Conference on Electronic Voting*, E-Vote-ID, pages 84–109, 2017.

[10] R. Bernstein, A. Chadha, and R. Montjoy. Overreporting voting: Why it happens and why it matters. *Public Opinion Quarterly*, 65(1):22–44, 2001.

[11] D. Bowen et al. Top-to-bottom review of voting machines certified for use in California. Technical report, California Secretary of State, 2007. https://www.sos.ca.gov/elections/ovsta/frequently-requested-information/top-bottom-review/.

[12] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri. Bridging the gap in computer security warnings: A mental model approach. *IEEE Security & Privacy*, 9(2):18–26, Mar. 2011.

[13] C. Bravo-Lillo, S. Komanduri, L. F. Cranor, R. W. Reeder, M. Sleeper, J. Downs, and S. Schechter. Your attention please: Designing security-decision UIs to make genuine risks harder to ignore. In *9th Symposium on Usable Privacy and Security*, SOUPS, 2013.

[14] M. D. Byrne and S. Bovair. A working memory model of a common procedural error. *Cognitive Science*, 21(1):31–61, 1997.

[15] B. A. Campbell and M. D. Byrne. Now do voters notice review screen anomalies? A look at voting system usability. In *USENIX Electronic Voting Technology Workshop/Workshop on Trustworthy Elections*, EVT/WOTE, 2009.

[16] Clear Ballot Group. ClearAccess administrators guide, 2015. https://www.sos.state.co.us/pubs/elections/VotingSystems/systemsDocumentation/ClearBallot/ClearAccess/ClearAccessAdministratorsGuideRev4-0-r0.pdf.

[17] A. Cordova, L. Howard, and L. Norden. Voting machine security: Where we stand a few months before the New Hampshire primary. Brennan Center, 2019. https://www.brennancenter.org/analysis/voting-machine-security-where-we-stand-six-months-new-hampshire-primary.

[18] L. F. Cranor. A framework for reasoning about the human in the loop. In *1st Conference on Usability, Psychology, and Security*, UPSEC. USENIX, 2008.

[19] M. De Jong, J. Van Hoof, and J. Gosselt. Voters' perceptions of voting technology: Paper ballots versus voting machine with and without paper audit trail. *Social Science Computer Review*, 26(4):399–410, 2008.

[20] R. DeMillo, R. Kadel, and M. Marks. What voters are asked to verify affects ballot verification: A quantitative analysis of voters' memories of their ballots, 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3292208.

[21] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: An empirical study of the effectiveness of web browser phishing warnings. In *26th ACM Conference on Human Factors in Computing Systems*, CHI, pages 1065–1074, 2008.

[22] S. P. Everett, K. K. Greene, M. D. Byrne, D. S. Wallach, K. Derr, D. Sandler, and T. Torous. Electronic voting machines versus traditional methods: improved preference, similar performance. In *26th ACM Conference on Human Factors in Computing Systems*, CHI, pages 883–892, 2008.

[23] A. J. Feldman, J. A. Halderman, and E. W. Felten. Security analysis of the Diebold AccuVote-TS voting machine. In *USENIX Electronic Voting Technology Workshop*, EVT, 2007.

[24] M. Fitts. SC chooses new voting machines that will print paper ballots but some fear it's not safe. The Post and Courier, June 10, 2019. https://www.postandcourier.com/article_f86632ce-8b83-11e9-8dab-5fb7858906cc.html.

[25] S. Fowler. Georgia awards new voting machine contract to Dominion Voting Systems. Georgia Public Broadcasting, July 29, 2019. https://www.gpbnews.org/post/georgia-awards-new-voting-machine-contract-dominion-voting-systems.

[26] S. N. Goggin and M. D. Byrne. An examination of the auditability of voter verified paper audit trail (VVPAT) ballots. In *USENIX Electronic Voting Technology Workshop*, EVT, 2007.

[27] K. K. Greene, M. D. Byrne, and S. P. Everett. A comparison of usability between voting methods. In *USENIX Electronic Voting Technology Workshop*, EVT, 2006.

[28] Indiana Fiscal Policy Institute. Vote centers and election costs: A study of the fiscal impact of vote centers in Indiana, 2010. https://www.in.gov/sos/elections/files/IFPI_Vote_Centers_and_Election_Costs_Report.pdf.

[29] D. Jones and B. Simons. *Broken Ballots: Will Your Vote Count?* CSLI Publications, 2012.

[30] D. Kasdan. Early voting: What works. https://www.brennancenter.org/sites/default/files/publications/VotingReport_Web.pdf.

[31] T. Kohno, A. Stubblefield, A. D. Rubin, and D. S. Wallach. Analysis of an electronic voting system. In *25th IEEE Symposium on Security and Privacy*, 2004.

[32] M. Lindeman and P. Stark. A gentle introduction to risk-limiting audits. *IEEE Security & Privacy*, 10:42–49, 2012.

[33] J. Mack. Who votes in Michigan? A demographic breakdown. MLive, 2018. https://www.mlive.com/news/erry-2018/11/340b0f9c406363/who-votes-in-michigan-a-demogr.html.

[34] S. Maneki and B. Jackson. Re: Comments on Ballot Marking Devices usage for the 2018 elections, 2017. Letter to Maryland State Board of Elections, citing SBE data.

[35] P. McDaniel, M. Blaze, and G. Vigna. EVEREST: Evaluation and validation of election-related equipment, standards and testing. Technical report, Ohio Secretary of State, 2007. https://www.eac.gov/assets/1/28/EVEREST.pdf.

[36] National Academies of Sciences, Engineering, and Medicine. *Securing the Vote: Protecting American Democracy*. The National Academies Press, Washington, DC, 2018.

[37] National Conference of State Legislatures. Funding elections technology, 2019. https://www.ncsl.org/research/elections-and-campaigns/funding-election-technology.aspx.

[38] R. G. Niemi and P. S. Herrnson. Beyond the butterfly: The complexity of U.S. ballots. *Perspectives on Politics*, 1(2):317–326, 2003.

[39] M. M. Olembo and M. Volkamer. E-voting system usability: Lessons for interface design, user studies, and usability criteria. In *Human-Centered System Design for Electronic Governance*, pages 172–201. IGI Global, 2013.

[40] S. Patil, R. Hoyle, R. Schlegel, A. Kapadia, and A. J. Lee. Interrupt now or inform later? Comparing immediate and delayed privacy feedback. In *33rd ACM Conference on Human Factors in Computing Systems*, CHI, pages 1415–1418, 2015.

[41] J. Petelka, Y. Zou, and F. Schaub. Put your warning where your link is: Improving and evaluating email phishing warnings. In *37th ACM Conference on Human Factors in Computing Systems*, CHI, 2019.

[42] Pew Charitable Trusts. Colorado voting reforms: Early results. https://www.pewtrusts.org/ /media/assets/2016/03/coloradovotingreformsearlyresults.pdf, 2016.

[43] Pro V&V. Test report for EAC 2005 VVSG certification testing: Clear-Ballot Group ClearVote 1.4 voting system, 2017. https://www.eac.gov/file.aspx?A=kOBM5qPeI8KZlJyADXYTieiXLwsxw4gYKIVroEkEBMo%3D.

[44] W. Quesenbery. Ballot marking devices make voting universal. Center for Civic Design, 2019. https://civicdesign.org/ballot-marking-devices-make-voting-universal/.

[45] W. Quesenbery, J. Cugini, D. Chisnell, B. Killam, and G. Reddish. Letter to the editor: Comments on "A methodology for testing voting systems". *Journal of Usability Studies*, 2(2):96–98, 2007.

[46] R. W. Reeder, A. P. Felt, S. Consolvo, N. Malkin, C. Thompson, and S. Egelman. An experience sampling study of user reactions to browser warnings in the field. In *36th ACM Conference on Human Factors in Computing Systems*, CHI, 2018.

[47] R. Rivest. On the notion of 'software independence' in voting systems. *Philos. Trans. Royal Soc. A*, 366(1881):3759–3767, October 2008.

[48] T. Selker, E. Rosenzweig, and A. Pandolfo. A methodology for testing voting systems. *Journal of Usability Studies*, 2(1):7–21, 2006.

[49] T. Selker, E. Rosenzweig, and A. Pandolfo. Reply to comment on: The Methodology for Testing Voting Systems by Whitney Quesenbery, John Cugini, Dana Chisnell, Bill Killam, and Ginny Redish. *Journal of Usability Studies*, 2(2):99–101, 2007.

[50] P. Stark. Conservative statistical post-election audits. *Annals of Applied Statistics*, 2(2):550–581, 2008.

[51] P. B. Stark. There is no reliable way to detect hacked ballot-marking devices, 2019. https://arxiv.org/abs/1908.08144.

[52] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying wolf: An empirical study of SSL warning effectiveness. In *18th USENIX Security Symposium*, pages 399–416, 2009.

[53] United States Senate Select Committee on Intelligence. Report on Russian active measures campaigns and interference in the 2016 U.S. election, 2019. https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume1.pdf.

[54] U.S. Census Bureau. QuickFacts: Ann Arbor, 2019. https://www.census.gov/quickfacts/annarborcitymichigan.

[55] U.S. Election Assistance Commission. Designing polling place materials. https://www.eac.gov/election-officials/designing-polling-place-materials/.

[56] Verified Voting. Ballot marking devices. https://www.verifiedvoting.org/ballot-marking-devices/.

[57] Verified Voting. The Verifier: Polling place equipment. https://www.verifiedvoting.org/verifier/.

[58] VSAP. Voting system for all people. https://vsap.lavote.net/.

[59] Wall Street Journal. Election 2018: How we voted in the 2018 midterms, November 6, 2018. https://www.wsj.com/graphics/election-2018-votecast-poll/.

[60] D. S. Wallach. On the security of ballot marking devices, 2019. https://arxiv.org/abs/1908.01897.

[61] M. S. Wogalter. Communication-human information processing (C-HIP) model. In M. S. Wogalter, editor, *Handbook of Warnings*, chapter 5, pages 51–61. Lawrence Erlbaum Associates, Mahwah, NJ, 2006.

[62] M. S. Wogalter and K. R. Laughery. Warning! sign and label effectiveness. *Current Directions in Psychological Science*, 5(2):33–37, 1996.

Appendix A
Poll Worker Script

Our poll workers followed four versions of the script below: a baseline version, and three variants that each add one line.

Variant 1: Before the voter begins using the BMD, a poll worker asks them to check their ballot before it is scanned.

Variant 2: Before the voter deposits the ballot, a poll worker informs them that it is the official record of the vote.

Variant 3: Before the voter deposits the ballot, a poll worker asks whether they have carefully reviewed each selection.

When Subject Arrives (Poll Worker A)

*Hello! Before you begin, please fill out this Institutional Review Board consent form. [*Point to form and pen.*] If you have any questions, feel free to ask.*

*You are about to participate in a study about the usability of a new type of voting machine. You will be using one of these voting machines to make selections on your ballot, which will be a truncated version of the Ann Arbor 2018 midterm ballot. Once you are finished, your ballot will be printed from the printer beneath the machine, and you can review your ballot and deposit it in the ballot box over there. [*Point out ballot box.*] Feel free to vote your political preference or not; no identifying information will be collected that could match you with your votes. If you would like to quit at any time during the study, just say so.*

Variant 1: *Please remember to check your ballot carefully before depositing it into the scanner.*

You may begin at any time.

Before Subject Deposits Ballot (Poll Worker B)

Variant 2: *Please keep in mind that the paper ballot is the official record of your vote.*

Variant 3: *Have you carefully reviewed each selection on your printed ballot?*

After Subject Deposits Ballot (Poll Worker B)

*Thank you for participating! You are now finished with the study, and should fill out the exit survey. [*Point to debrief survey computers.*]*

After Subject Completes Exit Survey (Poll Worker B)

*Thank you for your participation! You are now finished. If you have any questions about this study, you may ask them now, although I am unable to answer some questions due to the nature of the research. Here is a debrief form. [*Hand subject a debrief form.*] If you think of anything after you leave, you can reach [me/the principle investigators] through the information on the debrief form.*

If you know anyone who might like to participate, please refer them here; we will be here [remaining time].

Thank you again for participating!

Appendix B
Recruitment Script

An investigator used the following script to recruit library patrons to participate in the study:

Hello, do you have 10 minutes to participate in a study about a new kind of voting machine that is used in elections across the United States? This study will consist of voting using our voting machine and depositing a printed paper ballot into a ballot box, and then filling out a survey about the experience. If you would like to participate, we will need you to first sign a consent form. We will provide a flyer at the end of your participation with information about the study. We cannot make all details available at this time, but full details and research results will be made available within six months of the conclusion of this study. We thank you for your consideration and hope you choose to participate!

Appendix C
Slate of Candidates for Directed Voting Condition

We randomly generated a slate of candidates and provided a printed copy to voters in certain experiments. The handout voters received is reproduced below:

Race	Candidate(s)
Governor and Lieutenant Governor	Bill Gelineau and Angelique Chaiser Thomas
Secretary of State	Mary Treder Lang
Attorney General	Lisa Lane Gioia
United States Senator	Debbie Stabenow
Representative in Congress 12th District	Jeff Jones
Member of State Board of Education (Vote for 2)	Tiffany Tilley Mary Anne Hering
Regent of the University of Michigan (Vote for 2)	Jordan Acker Joe Sanger
Trustee of Michigan State University (Vote for 2)	Mike Miller Bruce Campbell
Justice of the Supreme Court (Vote for 2)	Megan Kathleen Cavanagh Kerry Lee Morgan
Judge of Court of Appeals 3rd District Incumbent Position (Vote for 2)	Jane Marie Beckering Douglas B. Shapiro
Judge of Circuit Court 22nd Circuit Incumbent Position (Vote for 2)	Timothy Patrick Connors Carol Kuhnke
Judge of Probate Court Incumbent Position	Darlene A. O'Brien
Judge of District Court 14A District Incumbent Position	Thomas B. Bourque